

Understanding begins at home

A British committee has suggested how the general understanding of science might be improved. In Britain, education cries out for revolution. But everywhere, science needs a more open manner.

THE appearance of the Royal Society's report (see p. 104) on the public understanding of science, and the existence of the committee that has produced it, is a mark of the general depression of the morale of the scientific community in Britain two years ago (when the committee came into being). For then it was generally believed that science and research would not have been so badly treated by a seemingly uncaring government if they were not also unfairly short-changed in public esteem. The most obvious dangers were that the analysis of an important problem would be biased by speculations about the most efficient way of creating a more effective lobby for the research enterprise or by inward-looking complaints about the failure of newspapers and other channels of communication to do the same job on a voluntary basis. In the event, the Bodmer committee has avoided these pitfalls. Instead, it has produced a valuable and liberal document that will help in two specific ways: by further strengthening the opinion that the most public ignorance of science (especially among members of the British government and their servants) can be traced to the educational system, and by creating a climate in which researchers are more willing to talk about their work. The sad thing is that a report along these lines should now be necessary in a community previously outstanding for the vigour with which science was a part of the general culture. Josiah Wedgwood's Lunar Society in the late eighteenth century was only one of the informal institutions that kept this spirit alive for the best part of 200 years.

What went wrong? The Bodmer committee is right in almost everything it says, but its analysis is over-timorous and, as such, may also be over-flattering to the scientific community everywhere, not just in Britain. Throughout Europe, the Enlightenment that occupied the closing decades of the eighteenth century was a conspiracy between intellectuals of all kinds, for whom rationality and optimism were the common unifying cement. The period spawned revolutions (in the United States as well as France), public educational systems (as in what became Germany), a flood of endearing but enduring literature, the techniques of business (such as joint stock-banks) and the technology that became the industrial revolution. That it should have been a period when people without specialist knowledge (of which there was not much) chose to take a lively interest in the natural world cannot have been an accident. In retrospect it appears that the circumstances were right for just such a revival of general interest in what science is about four decades or so ago, with the ending of the Second World War and the recognition of how much research had contributed to that end. (Awful though the war had been, it was clear that it could have been disastrous, so that there were also grounds for optimism.) Yet within a decade, in Britain, C.P. Snow found it necessary to write his tract on the "two cultures", a complaint that society had chosen to rob itself of profit and intellectual excitement by arranging that there should be two distinct, sometimes conflicting, intellectual camps.

The Bodmer committee is right to single out the British school system as one of the causes of this state of affairs in Britain, but the problem goes deeper. One of the reasons why people were more willing to argue about intellectual matters two centuries ago is that the divisions between ordinary people and the professional gatherers of knowledge were not then as sharp as they have since become. That there has since accumulated a vast bulk

of knowledge with which most non-academic people are unacquainted is a fact, but need not so often be an impenetrable barrier to general discussion. And the root cause of that development is the convention of self-certitude that has been taken up by academics, both in relation to students and, more alarmingly, among each other. In one respect, the advantages are obvious, in the other, the benefits are that people are politely restrained from too open mutual competition.

Among academics, scientists have earned themselves the reputation of being the least discursive, with consequences that are thoroughly unwelcome. The convention that one should not be too enthusiastic about one's own preoccupation is not merely a device for not giving good ideas away, or for avoiding the embarrassment of having them shown up as false, but it is a restraint on mutual stimulation. The convention that one should also be seen to be a deal smarter than one's students may be natural enough, but can hardly be necessary to the self-esteem of able academics and is in any case untenable, given the smartness of students nowadays. Yet the convention robs students of what should be essential stimulation. The frontiers of knowledge are also the boundaries of ignorance, and people whose teaching steers clear of the unknown provide their students with an ill-drawn map. Moreover, while the Bodmer committee has been understandably concerned in its analysis with what the "public" may know about science, it is more than likely that the general interest in this field is with the other side of this coin, what remains unknown.

That is one reason why the Bodmer prescription for the improvement of public understanding in Britain, good as it is, should be given an even more vigorous trial where it matters most, in the lecture halls and seminar rooms of the universities. If scientists, following Bodmer, are to tell the world why their own work is interesting, and to guess where it may lead, should they not also tell their students? If teachers are made thereby to seem a little more fallible, may they not also be more valuable as teachers, partly by stimulating those who listen to them, partly in providing a model of open discussion that will eventually communicate itself more generally? To the rejoinder that this is what outstanding teachers have always done, the counter-rejoinder is to ask how many teachers now regard the isolation of the intellectual nub of a topic, rather than its description, as the objective of an encounter with students.

So the moral, for scientists everywhere, is that public understanding begins at home, in the laboratory and the lecture room. For most practising scientists, the growing army of people who profess to write about science will be an irrelevance (and even a continuing embarrassment, given their tendency to ask questions that cannot be answered conveniently, or prudently). The benefits of more understanding between scientists and with their students would not necessarily be as distant as implied by the time taken for the same students also to be part of the general population. For in the past three decades, since the rot set in, scientists have had only moderate success in the competition that ultimately matters most to them — that for bright potential students. The arts and humanities have consistently done better than their practical value would suggest, no doubt because students enjoy such studies and are stimulated by them. Is that not a cause worth winning? □



Equipment up for grabs

US universities share others' troubles over equipment, but have more options to play with.

THE US research universities have been complaining with increasing conviction, over the past few years, about the need for modern equipment. Last week, the Committee on Science and Technology of the House of Representatives, apparently sympathetic to the complaints, was presented with the result of a three-year study of the problem by a task force of the Association of American Universities (AAU) under Richard A. Zdanis, vice-provost of Johns Hopkins University. The Zdanis committee's case is necessarily somewhat general, based more on trends discernible in federal budgets and the like than on the more compelling anecdotal evidence showing how particular research programmes of acknowledged worth are being hampered. To make up for this, the report is filled with a string of eminently sensible suggestions.

Before too many tears are shed, it should be recognized that what has happened in the United States is not very different from what has happened elsewhere, in Western Europe in particular. (Japan is a different case). Although the US government's spending on basic research continues to grow, the pace of growth is smaller now than in the happy 1960s, with the result that prudent administrators, in the universities and at the grant-making agencies, prefer to spend money on people rather than on research equipment. But during the same period, the design of standard equipment has been improved, making equipment more productive of material or data (which is its purpose, after all) but also more costly. Similarly, entirely novel kinds of research equipment have made their appearance, making possible tasks that could not previously have been tackled in university laboratories. The number of places now equipped with analytical equipment controlled by microprocessors, and even with the machines for making semiconductor chips for microprocessors, is a sufficient proof of that. That is the essence of the problem described last week. What can be done about it?

Mercifully, for US universities, the importance of the problem seems already to have been seized by the agencies. While the provision of research equipment through the regular research-grant mechanisms has by no means shrunk to nothing, the National Science Foundation, the Department of Defense (DoD) and the Department of Energy between them have recently committed very substantial sums to make good what seems to be a growing shortage. The AAU study itself is a manifestation of this awareness (and has indeed been supported by a grant from the National Science Foundation). Inevitably, the committee pleads for more of the same, together with the removal of some of the present impediments to easy dealing between the agencies and the universities — the requirement (under DoD grants and contracts) that there should be a search in the Pentagon's storehouse of discarded equipment before an equivalent piece of hardware can be provided at government cost is a worthy casualty. What the AAU committee seems not to have appreciated is the danger that the present alarm about the deficiencies of equipment will become institutionalized, and separate from other kinds of research support, with the consequence that there will be the eyes of two needles through which each research project must pass. There is also among the committee's proposals a trusting faith in the innocence of the Office of Management and Budget, which will presumably have an opinion on the suggestion that universities that have acquired equipment out of their own funds or otherwise should be allowed to charge annual depreciation against research grants when these materialize.

This is the bread and butter of the AAU report, to which it must be hoped that the Congress will respond. The frosting on the cake, the required reading for provosts from now on, is the section on self-help. Inevitably, universities are enjoined to be self-conscious about the uncovenanted costs of expensive equipment (such as start-up and maintenance), not to mention the

need to keep advanced equipment performing efficiently by means of a sufficient force of technical help. It is no surprise that individual universities have been deterred from entering costly fields of research, which may not be bad as it seems. Does the interest of the United States require that each of more than fifty research universities should be able to make a start on molecular beam epitaxy at a cost of \$1 million each? (And who was the chemistry professor whose installation at an eastern seaboard university cost \$2 million, expensive even now?) More down to earth, the committee had some useful suggestions as to how people may be able to buy expensive equipment by running two consecutive grants together, or even by leasing equipment.

More adventurous university administrators will be even more enthused by the recommendations in the AAU report that universities should more deliberately explore the use of debt financing for the purchase of equipment. (The committee responsible notes with some surprise that it knows of no university computer centre financed by such means.) Here the trick is to use the status of universities as tax-exempt institutions to raise money at low rates of interest, recovering the cost of grants not yet awarded; the AAU committee is right to commend such schemes while warning of the risks (chiefly, that expected grants do not materialize). But it may matter more than the committee allows that researchers do not like this way of acquiring equipment. For the short term, there is probably more to be said for making sure that university administrators understand the tax benefits of equipment grants and gifts by private donors. (AAU provides worked examples.) And in the language of resource economics, it is sad that there are still so few institutions that have followed the University of Iowa in setting up (and keeping going) internal arrangements for sharing not merely equipment but the problems that go with its ownership. □

Protection from whom?

Somebody should take a serious look at space launching costs.

THE US government has returned from the summer holidays in a sombre frame of mind on external trade policy. The Congress has been conscious for several months of the pressure from its constituents to keep foreign imports out of the United States. Now the administration is making conciliatory noises, offering still vague external negotiations with exporters to the United States. President Reagan's calculation is no doubt that it would be better to make small concessions now than to veto whatever comes up from Congress only to find the veto overridden. What everybody overlooks is the certainty that general protection beneath the shadow of the US federal deficit will have disastrous consequences for world trade.

But is it not unfair that people in the United States should be out of work (although the numbers are dramatically reduced this month) because people elsewhere will work for lower wages, or because their governments subsidize the products that they make? Before the Congress takes this argument too seriously, it should listen carefully to what the head of Arianespace was saying last week about the effects of US export subsidies on the space launcher business. The issue has been well rehearsed in the report from the Office of Technology Assessment on space competition, published earlier this year. The present basis for charging for the use of the US shuttle is that even if the cargo bay were unfilled, the shuttle would still fly. The present cost of a full bay is \$80 million, but it is intended that the cost should rise by 1988 to an amount estimated at \$125 million (in 1984 money). M. d'Allest's complaint is that the shuttle's present costs are less than they should be, although he is also open to the complaint that Arianespace can offer customers more attractive financing terms than NASA is allowed to offer. At best, the kettle is abusing the pot. At worst, the US shuttle is being as heavily subsidized in its competition with the European rocket as are the whole of the products of the textile industries now threatened from the United States. □

Japanese industry

Rival ministries collaborate on new research centres

Tokyo

FROM the end of this month, Japanese industry will be able to turn to a new Centre for Enhancing Research in Fundamental Technologies for help in carrying out basic research. Surprisingly, the new centre is to be run jointly by two ministries noted for their rivalry: the Ministry of International Trade and Industry (MITI) and the Ministry of Posts and Telecommunications (MPT).

Originally, the idea of a centre for basic research was dreamt up by MITI and finance requested in the 1985 budget. But at almost the same time, MPT began to demand that some of the money from the privatization of the state telecommunications monopoly, Nippon Telegraph and Telephone, which had been controlled by the ministry, should be used to finance a whole series of new telecommunication research centres. Although the initial demands of MPT were clearly excessive and seemed mainly inspired by a desire to dominate telecommunications research the government decided that MPT should at least share a new centre for research with MITI.

That decision came as something as a surprise, for MITI and MPT have been wrestling for some time to try to gain ministerial responsibility for data networks. MITI claims that as networks are just "linked computers", they fall into its traditional sphere of responsibility for computers; MPT, on the other hand, sees data links as being not much different from telephone lines and thus clearly its responsibility.

Interministerial struggles of this kind are quite common in Japan and make something of a nonsense of the Western image of a unified "Japan Inc.". Whatever their differences, anyway, in this case the two ministries now share responsibility for the new centre. It is not actually a research facility but an administrative body intended to help private industry to carry out basic research.

First, the centre will ensure that access to government research facilities is given to private industry at low cost. The idea is that the centre will act as a go-between so that better cooperation can be achieved between government, industry and the universities; lack of effective cooperation has long been one of Japan's weak points, despite Western myths to the contrary. The 16 research laboratories of MPT and MITI, including such giants as the Electrotechnical Laboratory, are particularly likely to be involved in such schemes.

Second, the centre will make direct low-interest loans to companies launching

joint fundamental research projects. In the event that the research does not succeed, the interest on the loan will not have to be repaid at all.

Where the research cannot be expected to produce any benefits in the near future, financial help will be given to set up new joint companies.

Third, the centre will take over the administration of the "Japan trust". This is a new charitable foundation set up to provide funds for foreigners to come and do research in Japan. Its capital has been provided by large industrial companies

Eureka

West Germany commits itself

BONN expects to have "no problem in financing" the French-inspired Eureka project for European collaboration in the development and production of new high technology products, a West German research ministry spokesman said on Monday. The provision of DM 100 million (£25 million) by the research ministry next year is "definite", and the money will be spent on feasibility studies and project planning, while another DM 1,000 million (£250 million) in grants to industrial participants in Eureka over the next five years is "almost decided".

These developments have been greeted with delight by the French government, which so far has been alone in making a clear financial allocation to Eureka. Before the summer, President François Mitterrand said France would commit FF 1,000 million (£90 million) to Eureka projects in 1986. Britain, by contrast, has hinted that no new government money will be found for the scheme, and has spearheaded attempts to raise interest among European banks and financial institutions.

But Bonn's interest is not free from strings. The West German ministry says that industry should provide the "bulk" of the funds (on top of the government's DM 1,100 million) that may be needed to develop products based on the Eureka technology, from robot tractors to supercomputers. And that "bulk" may be very large.

The inspiration for Eureka was the potential industrial spin-off from the United States Strategic Defense Initiative (SDI), on which \$26,000 million (£20,000 million) is expected to be spent over the first five years, and Europe seems bound to make a comparable investment if there is to be a comparable effect. The publicly announced European commitments to

and it was founded partly to avert criticism that Japan sends many researchers to learn advanced techniques in the United States but offers very few opportunities to foreigners in return. Researchers to be invited under the scheme will be selected by the centre and responsibility taken for negotiation over rights to patents stemming from such international research.

Industry's recent spending spree on new basic research laboratories (see *Nature* 311, 404; 1984) shows that enthusiasm for basic research exists. The ¥4,000 million (US\$17 million) available to the centre for this year's costs should certainly help to make the centre's services popular. But the traffic is not all one way; increased links with industry will also help to ease government research institutes' worries that they could be left behind by the increasingly vigorous laboratories attached to large companies.

Alun Anderson

Eureka, however, amount to hardly 2 per cent of this sum.

Nevertheless, the West German commitment is symbolic and crucial, both for the "special relationship" between France and Germany and for Eureka. Relationships have been eroding badly in recent months, notably in Chancellor Helmut Kohl's commitment to SDI and President Mitterrand's rejection of it. But, meanwhile, German negotiators in Washington are said to have become somewhat disenchanted with the US position, and, in particular, concerned that technology developed under the SDI umbrella with European help might be held under such strict export controls by the United States that its application in ordinary commerce might prove impossible. Hence the German position is felt to be shifting from an initial enthusiasm for SDI towards greater interest in the commercially-oriented Eureka.

On Eureka, apart from financial questions, they say in Bonn that "we still need to know what the Eureka project will be", and there is detectable concern that despite the continuing work of a 15-nation (plus the European Commission) expert group, there is no certain agreement on exactly what Eureka will do.

In France, there has been interest in supercomputers, in the use of lasers for surface treatments, the applications of a small intense electron synchrotron (also of interest to Italy, which seeks to place such a device in Trieste, as compensation for losing the European Synchrotron Radiation Facility to Grenoble), robotics and genetic engineering of drugs; Germany is also interested in environmental technologies. And with such a wide variety of issues available, it is still an open question whether anything concrete can be decided.

Robert Walgate

AIDS

Poland's minister for prophylaxis

THE Polish Ministry of Health and Social Welfare, has appointed a special "plenipotentiary" for AIDS (acquired immune deficiency syndrome), it was announced last week. Professor Jerzy Bonczak, who is already a deputy minister, will be in charge of a programme of epidemiological examination among those considered to be particularly at risk: homosexuals, bisexuals, drug addicts and haemophiliacs. This will be followed by a nationwide screening of blood donors. Clinicians are being trained, and a contingency treatment centre prepared. A special grant of hard currency has been made to the Ministry of Health and Social Welfare to enable rapid diagnostic kits to be purchased from the West.

The threat of AIDS has been a matter of public alarm in Poland for the past two years. The run-down of the pharmaceutical and medical supply industries under the Gierek regime meant that, by 1980, even such basic equipment as catheters and hypodermics were virtually unobtainable. Lech Walesa's "Medical Rescue Bank", which arranged supplies from Western charitable organizations, has luckily survived the overthrow of Solidarity under the patronage of the Roman Catholic Church, but throughout 1982 and 1983, "disposables" were in such short supply that hospitals had no option but to sterilize and reuse such items over and over again. Although supply has improved, according to one Warsaw doctor, some reuse of disposables is still necessary.

Drug addiction, moreover, is a growing problem in Poland. Poppies are a cash crop (the seeds are used for cooking), and the illegal manufacture of opium derivatives from the straw has been going on for decades. But until 1980, the government refused to acknowledge the problem, and consequently took no counter-measures. Now, there are estimated to be at least 200,000 addicts, two-thirds of whom are under 21.

So far, according to Jan Suchowiak, director of the Department of Public Health of the Ministry of Health and Social Welfare, Poland has had no "clinically diagnosed" cases of AIDS, although there have been "a few suspect cases", in which the "probability of AIDS" was later "excluded".

In general, details on AIDS within the socialist bloc are hard to come by, partly because of the tendency in the press to play down information liable to cause alarm and despondency but also, undoubtedly, due to the initial identification of AIDS as a disease of homosexuals. In the Soviet Union, where active homosexuality is a legal offence for which compulsory psychiatric treatment may be ordered by the courts, AIDS was at first described as a "disease of capitalism", thus simul-

taneously castigating the West and reassuring the Soviet public. The same approach was abandoned in Czechoslovakia, when two suspected cases of AIDS (a visitor from Africa and one of his contacts) were reported in Prague.

So far, only Poland has announced plans for dealing with a possible outbreak of AIDS, but there is growing concern with what may be styled "AIDS-related" issues in other socialist countries. Thus at the end of June, Budapest radio reported a "violent debate", at a session of the Cen-

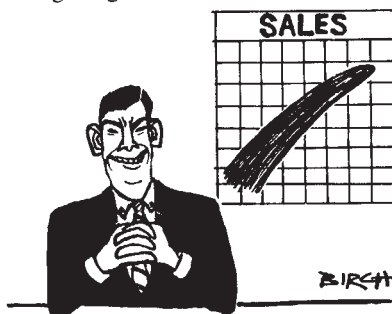
Halley's comet

Per astra ad pecuniam

Tokyo

AFTER the successful launch of a probe to Halley's comet last month by Japan's Institute of Space and Astronautical Science (ISAS) (see *Nature* 316, 669; 1985), preparations are in hand for the arrival of the comet later this year. And, if all goes according to the plans of Japan's Halley's Comet Society, the comet's nearest approach in March/April next year will culminate in a blackout of Tokyo and an invasion of Australia by three jumbo jet-loads of Japanese.

In a country where short-term fads can make millions and where getting drunk while gazing at the Moon is a time-



honoured tradition, Japan's advertising giant Dentsu was quick to realize the potential of the return of Halley's comet. It was at Dentsu's suggestion that the Japan Halley's Comet Society was founded in March last year. Headed by prominent academics, including Sadao Murayama of the National Science Museum, Shinya Obi, emeritus professor of Tokyo University and Tatsuzo Obayashi, professor of ISAS, the society has the lofty aim of "increasing the Japanese people's awareness of natural sciences and astronomy", according to spokesman Noriaki Funada. The society boasts a membership of over 6,000, ranging in age from three to ninety.

The society publishes a brochure on Halley's comet four times a year, is holding meetings at planetariums in cities all over Japan and is organizing observing meetings in various parts of the country

tral Leadership of the Union of Medical and Health Workers, over the lack of proper sterilization facilities in hospitals and dental surgeries. Also in June, an unusually sympathetic "academic meeting" was held in Leipzig on the "psycho-social problems of homosexuality", with particular emphasis on counselling. At least one country, however, is still prepared to tackle the AIDS problem by a straight denial; two weeks ago, the Yugoslav Federal Secretariat for Information held a special press conference to dispel public concern by announcing that, despite rumours to the contrary, not a single case of AIDS had been "registered" in Yugoslavia. **Vera Rich**

between autumn and spring next year. But the society's two big events are a plan to turn the lights out in Tokyo during the early hours of 21 March and a trip to Australia in April next year.

The "blackout campaign" follows a similar attempt to dim the lights of Japan thirteen years ago, when the Giacobini-Zinner meteorite shower was expected. That campaign was only a partial success — although big companies like Matsushita Electric were persuaded to switch out their lights, the meteorites failed to turn up. A similar problem may bedevil next year's blackout; on 21 March Halley's comet will be low in the sky, skirting Tokyo's skyscrapers at 15° elevation and will have a magnitude of only 4.4, barely visible to the naked eye. Society officials are undaunted, however, and consider the campaign will be a success if it gets the Japanese public to look heavenward.

For a chance to see Halley's comet high in the sky, the society is chartering three jumbo jets from Japan Air Lines (JAL) to go to Bathurst, New South Wales, the following month. About half the seats have already been filled since places went on sale last month, despite the Y378,000 (£1,150) price. Akira Fujii, a prominent amateur astronomer, is making a 51-cm reflecting telescope for the society and Japan's TBS television station plans to broadcast pictures taken through it live from Bathurst via a satellite link-up.

Following the recent unbridled success of an advertising campaign associated with the import of koalas to Japan, in which the cute cuddly bears were used to sell everything from key holders to piggy banks, the sponsors of the Japan Halley's Comet Society, including JAL, Kanebo, Minolta and Fuji Film, will no doubt take full advantage of the publicity associated with the arrival of Halley's comet. In fact, Kanebo is already using the society's comet-like logo to sell cosmetics, while telescope manufacturers report a doubling in telescope sales this year.

David Swinbanks

US science education

Illinois plans elite high school

Chicago

THE state of Illinois, as part of its recently approved Education Reform Bill, has set aside \$500,000 to plan a special school for gifted children from 10th to 12th grade (16 to 18 years). This Science Academy is largely the idea of Leon Lederman, director of the Fermi National Accelerator Laboratory, and it will be located near Fermilab, west of Chicago, but will take in about 800 residential students from the whole state. Apart from the benefits for those who may attend the school, Lederman sees the academy as a morale booster for the whole public school system in the United States.

This development comes after a period when the quality of public education, especially in mathematics and the sciences, has become almost a political issue. Research institutes and graduate schools are seeing increasing numbers of foreign students, while US students are turning towards potentially more lucrative careers in law and business. The causes of this general disenchantment with science are no doubt complicated, but much of the blame is being put on meagre and unexciting school science curricula, taught by underqualified staff. Although President Ronald Reagan and others have from time to time bemoaned the poor state of US high schools, extra funds to remedy the problems have not appeared, on the familiar pattern that the quality of education can be improved by exhorting teachers to greater effort, without troubling to pay them better or buy them more equipment.

The new academy will be designed and overseen by trustees appointed by Illinois, but a study group of teachers and researchers from the Chicago area has already produced a model curriculum. The aim is to offer a complete education in arts as well as sciences, both through coursework and with students working individually under the supervision of teachers, or visitors from local universities and high-tech companies. Lederman hopes the academy will allow gifted children to study and learn in the company of their peers. In Illinois, with many small rural schools, isolation and its consequent frustration can be the biggest hindrance to the development of exceptional pupils.

Although the study group first embarked on the idea of separate institutions for the sciences, the arts, the social sciences and so on, the advocates of these schools all arrived at similar curricula embodying plans for a comprehensive liberal-arts education. The upshot is that the Illinois academy should, in the long term, cater for the needs of all gifted children, in all subjects.

One of the Illinois study group is Mr F. Borden Mace, principal of the North Carolina School of Science and Mathemat-

ics, an advanced, two-year (11th and 12th grades) high school which has been running for five years. Mace's experience, in both shaping and running the North Carolina School, seems powerfully to have influenced the Illinois proposal. The academy will go beyond its precursor in providing an accelerated education equivalent to a first year in college, enabling the very best students to join second year university courses directly.

Another of the Illinois academy's antecedents, and the origin of much of Lederman's interest in science education, is an informal programme called Saturday morning physics, which brings local high-school students to Fermilab for lectures and demonstrations by staff scientists. Lederman says he has found in these students an enthusiasm for science which is stifled by lack of interest and excitement at their schools. He blames this condition mostly on the school administrators who, since the revitalization of the education system after the shock of Sputnik in 1967, have since lapsed into complacency.

Concern over the perceived superiority of Japan in the application of research in industrial needs has now led politicians to worry again about the erosion of US science and technology; many see the re-establishment of a sound science education as a first step in the cure.

Even so, the plan for the Illinois academy has not been trouble-free. Some Illinois politicians do not wish to spend scarce education dollars on a small fraction of the school population; the study group estimates, however, that the annual cost will amount to about \$6 million out of a state education budget of \$2,000 million, or a running cost per student only three or four times the state average.

Local industries, especially some of the high-tech companies newly arrived in the Chicago suburbs, will also contribute funds in the expectation of better access to local universities and to the students themselves.

A second objection to the plans was the accusation of elitism. To Lederman, the word is not necessarily pejorative. He says that while the academy will be of obvious benefit to a small number of students, there will be a "trickle-down" effect that will improve the quality of education throughout Illinois.

Exceptional children will have teaching equal to their needs, and schools across the state will, he hopes, take the same pride in sending their best pupils to the academy as they would in sending them to an Ivy League college.

Teachers too will be able to take summer courses to increase their knowledge and teaching abilities, so spreading the advantages of the academy throughout the state.

David Lindley

Japanese blood

Supply edges nearer demand

Tokyo

FOR the past ten years, Japan's insatiable demand for imported blood has been under criticism from the World Health Organization (WHO). Now it looks as though things will change, with plans announced this week to bleed Japan's own citizens a little more heavily.

Japan has been criticized because it has been estimated that the country buys up between 25 and 32 per cent of the world supply of blood plasma, pushing up prices and making it difficult for less developed countries, many of which face difficulties in organizing blood donation campaigns, to buy plasma. WHO has several times asked Japan to try to become self-sufficient in blood supplies. This has been achieved for whole blood but not for plasma or blood products.

The problem is not that Japan is short of blood donors; in fact it has more donors per head of population than anywhere else. But only 200 ml of blood is taken at each donation, a little less than half that routinely taken in the United States and Europe. This week, however, the Ministry of Health and Welfare announced that from next spring, 400 ml of blood will be taken from each donor. At the same time, the age limit for donors is to be raised from 16 to 18 and the minimum body weight to 50 kg (110 lb, 7 stone 12 lb). Increasing use will also be made of techniques to take only blood plasma and platelets from the donor while returning red blood cells. With the new scheme, dependence on foreign blood plasma should fall to 50 per cent from the present 85 per cent within a few years.

Measures of this kind have been under discussion for some time but it seems that cases of AIDS (acquired immune deficiency syndrome) among haemophiliacs, including the death of two Japanese haemophiliacs from AIDS, have spurred action. At present almost all the imported plasma comes from the United States.

This week, the *Yomiuri* newspaper revealed that a large shipment of factor IX, used in the treatment of haemophilia B, had come from a US company that discovered that one of its donors had developed AIDS. By the time this information reached Japan, however, the shipment had been used up at hospitals.

Critics of the plan, while welcoming increasing self-sufficiency in plasma supplies, point out that insufficient efforts are being made to reduce demand. Use of plasma and blood products is unusually high in Japan. But that is part of a much wider problem of the private medical system: doctors can earn more if they prescribe more — and more expensive — drugs.

Alun Anderson

US tax reform

Colleges alarmed by new plans

Washington

US UNIVERSITIES are still up in arms about President Reagan's proposal to overhaul the tax system, even though the features of the original plan that caused most apprehension have been toned down. While the President is taking every opportunity to rally support for his proposal, the American Council on Education (ACE), an organization representing a wide spectrum of higher education institutions, has

them, but ACE claims that the substitute proposals are not much better. "What they've given with one hand they've taken away with the other", says one ACE analyst.

ACE's principal worry now concerns a special tax provision called "minimum tax", to be used exclusively for wealthy people who have their tax affairs so well arranged that, without the minimum tax provision, they would pay no tax at all. Under existing tax arrangements, charitable donations are not counted in with the other deductions in calculating minimum tax, but the President's intention now is that they should be, increasing the cost of giving by the wealthy. ACE, quoting Lindsey again, says that 65 per cent of all donations of appreciated property come from taxpayers with an adjusted gross income of more than \$100,000 who use the minimum tax provision. The effect of the proposed change would be "devastating", he says.

ACE's concern about charitable donations is not confined to wealthy givers. A

different provision in the plan that would primarily affect lower-income taxpayers would also reduce contributions to non-profit organizations — by \$6,700 million, according to ACE's estimate. And other aspects of the reform proposal that would affect higher education have nothing to do with charity; they include a proposal that might lower state tax revenues, a proposal to tax scholarships as personal income and a restriction on the use of tax-exempt bonds issued by state and local governments that would severely restrict their use by institutions of higher education.

The future of the plan is hard to predict. Despite President Reagan's apparently firm commitment to reform, there is a marked lack of enthusiasm in Congress and indeed in the country generally, especially among those who might benefit most from the existing array of tax loopholes to suit every pocket. Speedy progress on tax reform would certainly be a feather in the cap for Congressman Dan Rostenkowski, chairman of the committee in the House of Representatives that will draft the legislation. But other demands on Congress's time, mean that the proposal is not going to get an easy ride.

Tim Beardsley

Remember, folks —
charity begins at home!



pulled out all the stops for its lobbying effort to oppose the plan in its present form.

The plan seeks to rationalize and simplify the US domestic tax system so as to avoid some of its more flagrant injustices while not changing significantly the total amount of revenue raised. One of its important elements would severely restrict the deductions against gross income that can be claimed for charitable donations, which at present allow taxpayers to decide whether to make payments to the Inland Revenue Service or to charities of their choice — a system open to abuse.

The effect of the change, according to ACE, will be to increase substantially the cost of giving to legitimate public-service institutions such as colleges and universities, and the council fears a precipitous drop in charitable donations as a result. ACE quotes a study by Lawrence Lindsey of Harvard University which concludes that the tax plan would reduce donations to non-profit organizations by \$11,000 million, or 17 per cent. Private universities alone receive \$3,000 million a year in charitable contributions.

Under an earlier version of the tax-reform plan, charitable contributions would have been deductible only if they exceeded 2 per cent of a taxpayer's gross income, and donations of non-cash items such as real estate and works of art would no longer have been deductible at their full market value. The outcry over these provisions led the President to abandon

US Arctic

Strategy and science mix

Washington

THE National Research Council is urging a major expansion of US Arctic research. In a report* published last week, the council identifies twelve "issues of national concern" in the Arctic, and lists the most urgent research needs. A new dedicated polar research ship and a permanent logistics base near the Arctic coast are thought to be the minimum requirements.

The Arctic Research and Policy Act, signed by President Reagan in July 1984, called for a comprehensive 5-year research plan for the Arctic. The present study, under the chairmanship of Charles Bentley of the University of Wisconsin, was undertaken at the request of the committee preparing the research plan. Unlike many reports from the National Research Council, this explicitly relates scientific research to US needs in, for example, defence, weather forecasting, energy and materials, transport and communications.

Bentley's committee did not attempt to decide priorities between scientific research areas, though within each of the twelve considered, it gave rankings by importance. Thus, in near-Earth space physics, it attaches most importance to improvements in predictions of auroral, ionospheric and magnetic perturbations, which can affect communications. In atmospheric sciences, oceanography and biology, the committee attaches most importance to the value of the relatively undisturbed Arctic environment for measur-

ing and monitoring long-term changes, for example in atmospheric trace gases and ocean currents. Much basic ecology also remains to be described.

The report says the United States should concentrate its efforts in Alaska, giving suitable regard to the wishes of local inhabitants, but also stresses the need for international cooperation in Arctic studies. The present "fragmentary" domestic organization of Arctic research needs to be improved, with "big science" projects sharing logistical support. More satellite ground stations should be built, to receive data from both US and foreign satellites, and data should be published in the open refereed literature.

High priority should also be given to geological maps and surveying for oil and minerals, according to the report. As a practical engineering matter, there is an urgent need for better understanding of the formation and distribution of permafrost, with an observation network to monitor changes. And there should be an integrated research programme on sea-ice, concentrating in particular on engineering to withstand ice forces and the use of snow and ice as engineering materials. But this will require, according to the committee, better educational opportunities and incentives to encourage careers in Arctic science and engineering.

Tim Beardsley

*National Issues and Research Priorities in the Arctic. Polar Research Board, National Research Council, 1985.

Nuclear business

British fuel company trims sails

BRITISH Nuclear Fuels Ltd (BNFL) plainly expects a continuing public relations battle over the operation of its reprocessing plant at Sellafield in Cumbria, so much is clear from BNFL's annual report published last week. Sellafield has been the centre of the debate over the company's safe use of nuclear technology since an "unscheduled discharge" of radioactive waste into the Irish Sea resulted in a £10,000 court fine in July 1984. More recently, a motion before the European Parliament has called for the temporary closure of the plant.

Expansion modulated by concern for safety is BNFL's new image of itself. Its export earnings, for example, have increased by more than 40 per cent, from £91 million in 1984 to £128 million in 1985. This business, resulting principally from uranium enrichment, has helped to raise turnover from £460 million to £545 million. The company also points to the million pounds a day it expects to spend on new developments over the next 10 years, mostly financed by internally generated funds and advances from customers.

The centrepiece of this expansion is the £1,300 million thermal oxide reprocessing plant (THORP) at Sellafield. The plant, which is planned to be onstream by 1990, is already booked to capacity for the first ten years of operation, with Japan an important customer. At least until the end of the century, according to BNFL, a fleet of ships will be needed just for transporting these wastes from Japan. BNFL officials emphasize that its contracts usually include options to return treated wastes, and the company is "effectively under instruction" to exercise these options in the mid-1990s.

BNFL's bottom line, however, shows the effects of the Windscale court case. Though operating profits also increased, to £138 million, pre-tax profit for the year 1985 was only £68 million, compared to £71 million in 1984. The company ascribes this stasis to production setbacks in nuclear fuel reprocessing and to the commitment of funds to the reduction of radioactive discharges from Sellafield. BNFL claims that these programmes of environmental protection have "undoubtedly influenced the public to take a more favourable attitude towards the company's reprocessing operations".

One of the safety-inspired plants at Sellafield is the £126 million Site Ion Exchange Effluent Plant (SIXEP), which uses clinoptilolite from the Mojave Desert in the United States as a raw material for the extraction of caesium and strontium ions. Data from the computerized plant, according to the company, are not yet conclusive on its effectiveness. Another new project is the construction of £150 million effluent treatment and storage plant for

liquid discharges at Sellafield. The goal is to cut discharge of long-lived alpha radiation emitters from 370 curies in 1984 to less than 20 curies a year. BNFL chairman Con Allday emphasizes that these levels of sea discharges are far lower than those justified by "any simple cost/benefit analysis".

BNFL is not in a position to take quite as positive an attitude toward nuclear waste as was expressed at the British Association of Science meeting at the end of August. There officials from the National Radiological Protection Board and UK Atomic Energy Authority pointed to the high standards of the nuclear industry, describing radioactive waste

disposal as a matter of public education rather than public danger. BNFL claims, however, that such incidents as the 1983 Sellafield discharge have had "no great effect" on the environment, regardless of how data are interpreted by groups of outside critics.

Two important issues in the public relations battle over Sellafield are job protection and secrecy. The Sellafield plant employs over 10,000 people, making BNFL the second largest employer in the area. Neither these employees nor the public at large, according to the British Labour Party, are capable of judging the environmental impact of Sellafield because of the secrecy surrounding government and company reviews of safety at the site. Labour, in other words, would support an "open review at places like Sellafield".

Elizabeth Collins

Animals in research

US rules to be made tighter

Washington

THE University of Pennsylvania's Head Injury Laboratory, whose research has been temporarily halted pending an investigation by the National Institutes of Health (NIH) (see *Nature* 25 July, p.286), now faces the prospect of longer term closure. The US Department of Agriculture, after looking at videotapes stolen by an animal rights group that broke into the laboratory last year, is charging the university under the Animal Welfare Act, specifically with respect to operations performed using inadequate anaesthesia, insanitary conditions and improper postoperative care. The university faces a £4,000 fine if found guilty.

Meanwhile, the Head Injury Laboratory last week replied to NIH's report, which found evidence that its regulations were being ignored. NIH will "evaluate" the reply and decide later this month whether to continue support of the research programme. But further pressure is being brought to bear by Congress, where two representatives will seek to block support for next year when the NIH appropriations bill is discussed, thus preempting any action NIH might take.

The events surrounding the University of Pennsylvania reflect the general problem of laboratory animal welfare in the United States, where legislation is less strict than in many other countries (see *Nature* 311, 295; 1984). NIH issue guidelines for the use of animals in laboratories that demand adequate care and surgery to be performed by "trained experienced personnel"; a committee containing at least one veterinarian must regulate research at each institution.

The latest modifications to the guide, imposing stricter conditions, come into force at the end of the year. There are also regulations for research supported by other organizations, such as the National

Research Council and the Public Health Service. But the Animal Welfare Act is long overdue for reform. A recent study by the Animal Welfare Institute found that, using the "most optimistic assumptions", one-quarter of registered research facilities are not meeting the act's requirements.

Previous attempts to amend the act have failed. But increased public awareness of the mistreatment of laboratory animals, typified by the case of the University of Pennsylvania, has resulted in new optimism that Representative George Brown and Senator Robert Dole

We were carrying out an Action Research Project on human morality....



will next month successfully introduce their bill to amend the act.

The bill will considerably strengthen the regulations, requiring training for researchers, the use of pain killers and justification of the use of experimental animals as well as better pre- and postoperative treatment. Fines for the infringements of the law will be increased and committees, including a veterinarian and an informed outsider, will be responsible for the treatment of animals at each institution where research is performed. The amendments are supported by the American Physiological Society.

Maxine Clarke

Soviet libraries

How Novosibirsk keeps up

SOVIET engineers and applied scientists consider the updating of their professional knowledge to be of "vital importance", according to a survey reported in *Pravda* two weeks ago. This, of itself, is fairly surprising. For the past ten years, the Soviet media, including the slogans of the May Day and Revolution Day parades, have urged them to keep abreast of the latest developments in their profession, while the universities, the higher education institutes, the radio "universities of the masses" and the All-Union "Znanie" (knowledge) Society maintain a vast network of updating and upgrading courses. Unexpectedly, however, the *Pravda* report indicated that few of those interviewed attend formal refresher courses or use specialized science libraries. This result is particularly surprising, because the survey was conducted in Novosibirsk, the home of the Siberian branch of the Soviet Academy of Sciences and of several major research universities, science libraries and information centres.

Participation in refresher courses brings a number of fringe benefits for Soviet scientists. There is generous provision for study and examination leave and, officially, a person enrolled on such a course cannot be compelled to work overtime. Yet out of 626 people interviewed, only 1.3 per cent are enrolled in university courses (leading to a degree) and 26.4 per cent in job-related "higher qualification courses". No fewer than 70.4 per cent said that they were studying on their own.

Pravda rightly points out that private study can be effective only if it has a proper information base. More than half of those interviewed claimed that they had too little time to use the Novosibirsk libraries as much as they wished. Often, they said, the time spent travelling to and from a library was several times that spent consulting the works they had gone to read. More than 70 per cent complained of the difficulty of finding what they wanted to read, even with the help of catalogues designed to facilitate searching. More than 80 per cent said they preferred to find the information they require directly at their work-place; about 30 per cent liked to study at home by way of television, radio, books or journals and only 14 per cent favoured going to libraries. When asked "from what source do you most often obtain information that is valuable to you?", the most popular answers were journals and television, with books in only seventh place.

The *Pravda* commentator, A. Fabrichnyi (a senior lecturer at the Novosibirsk Higher Party School), suggests that the figures indicate a need to rethink and improve the "scientific and technical propaganda system". The media, in general, should be involved more closely with the

dissemination of new scientific knowledge, particularly the high-circulation periodicals of individual industries, which, paradoxically, contain far less material of direct use to the industrial scientists than do the major all-union dailies and weeklies.

What the commentator fails to note is that the basis of the survey seems ill-defined, with improvement of qualifications covering both the sustained study required to obtain a higher degree and the reading of occasional information sheets from the Scientific and Technical Information Centre. The disproportion between travelling to a library and time spent in it, the emphasis on television and journals and the preference for working alone suggest that improving one's qualifications may often mean nothing more than browsing through professional journals or watching a science programme on television.

Vera Rich

Public understanding

Royal Society urges activity

THE Royal Society will soon be awarding an annual prize to the British scientist or organization doing most to further public understanding of science if it follows all the recommendations of the committee on the subject under Dr W.F. Bodmer (Imperial Cancer Research Fund) published this week*. The essential argument of the report is that public understanding is a public good which scientists have a "professional responsibility to promote".

The most subversive of the proposals in the report, which is said to have been accepted by the council of the Royal Society, are however those concerning the British educational system. Thus the Bodmer committee asks that no school student should be allowed to concentrate exclusively on science or non-science studies "even after the age of 16", and that students at universities should be helped to a broader education either by courses of general study or in some other way.

In an attack on present practices in British secondary schools, the Bodmer report says that the education of post-16 students is "appallingly narrow" and that the degree of specialization now common, entailing as it does a "wasteful repetition" at school of what is later studied at university, is "totally unnecessary". The document goes on to say that universities and other institutions of higher education help to restrict the broadening of education by their "unnecessary and unreasonable" demands for specialization and says that claims that reform would undermine the quality of British higher education are belied by the success of school systems elsewhere.

Research data

University sues laggard students

Washington

WASHINGTON State University has filed a \$150,000 lawsuit against two graduate students who, it alleges, have refused to turn over data on trace gases in the atmosphere that were gathered during an Antarctic research project. An affidavit signed by the principal investigator of the project says one of the students has threatened to destroy the data if the university tries to recover them.

The issue is important because it raises questions about the contractual obligations of people other than principal investigators to the grant-making agencies supporting university research. Questions about the responsibility of students to their supervisors, and of universities for their students, also arise.

The data were gathered as part of a four-year project supported by the

The committee also wants a broader education, including experience of work outside the educational system, for science teachers, and says that in-service training should be an integral part of their conditions of service, not "an optional extra for enthusiasts".

Broadly speaking, the Bodmer committee is appreciative of the treatment accorded to science in the British press but especially by the broadcasting media. It gives a sympathetic account of the problems encountered by science journalists in the internal competition for newspaper space, but also says that the low opinion among editors of the importance of science in newspapers is belittling and "unwarranted". The committee complains that, with one exception, it failed to evoke a response from the newspaper editors whom it asked for help.

On the scientific community's responsibility, the committee urges that scientists are not excused by the emergence of a substantial number of professional writers on the subject (with whom the committee also notes a persisting professional mistrust). Apart from the proposal that the Royal Society should award an annual prize, the committee suggests that PhD candidates might be asked to produce a summary of their work for a lay audience, that the Royal Society should take the initiative in organizing press briefings for science journalists and that it should set up a standing committee to monitor progress in the field. □

*The Public Understanding of Science (ISBN 0 05403 25756), from the Royal Society, 6 Carlton House Terrace, London SW1Y 5AG.

National Science Foundation (NSF). The students, Stephen and Annette Waylett, spent the 1983 Antarctic winter at Palmer Station, Antarctica, gathering data on atmospheric chlorofluorocarbons, ozone, carbon dioxide, methane and nitrous oxide, together with supporting meteorological observations. The data were to have been included in master's degree dissertations.

Stephen Waylett had already failed one essential course for his degree and had not completed another when the couple left, although both difficulties could have been rectified later, according to a former supervisor.

The couple did supply the university with weekly averages of measurements during the period of observations, so there seems little doubt that measurements were correctly made. The university contends that, as the intermediary which paid for the research, it owns the data. University officials say they have offered to let the couple keep a copy of the data and that the Wayletts' research contribution would be acknowledged in any future use of it.

NSF's legal department says it knows of no previous instance of a researcher refusing to release data gathered under a foundation grant. The grant conditions do not apparently specify who formally owns data gathered; that they will be made public is simply presumed.

NSF has said it will stay out of the dispute between the university and the Wayletts, at least until the final report from the \$600,000 research project falls due next summer. Although final reports are often fairly brief, the university will need the Wayletts' data, which represent year two of a four-year series, for statistical comparisons, according to the project's principal investigator.

Those who know the Wayletts are at a loss to account for their behaviour, although they did apparently have some complaints about their tuition. The couple's whereabouts, and that of the data, are now unknown. A temporary restraining order issued by a federal court judge forbidding the couple to destroy or damage the data has still not been served after two weeks.

Tim Beardsley

Framatome joins the fold

FRAMATOME, the French nuclear company which is the largest constructor of nuclear reactors outside the United States, has been taken over by a French nationalized group, the Compagnie Générale Électrique (CGE), a move that given the French state essentially total control of the nation's nuclear industry.

Framatome not only constructs the nuclear component of French pressurized water reactors (PWRs), but has an affiliate Novatome, the constructor of the French fast breeder Superphenix. CGE itself controls both Alsthom Atlantique which makes the turbogenerators for the reactors and Câbles de Lyon which provides the electrical cabling. Moreover, the civil engineers Dumez, involved in the construction of the buildings for many French power stations, is to become a minority (10 per cent) shareholder in Framatome, and the national atomic energy authority, the Commissariat à l'Énergie Atomique (CEA), will take 35 per cent. CGE will be the leading shareholder with 40 per cent, the national electricity utility Électricité de France (EDF) will take 10 per cent "to see the bills", and Framatome staff will have the remaining 5 per cent.

The reconstruction of Framatome was necessary because of the collapse last year of Creusot-Loire, the private French steel giant, which held 50 per cent of Framatome (CEA had the other 50 per cent), and which the French government refused to bail out. This led to long discussions over the future of Framatome, clouded by the end of major purchasing of PWRs by EDF. For EDF, Framatome expects to be constructing four nuclear systems (pressure vessels and associated steam generators)

next year, but only two in 1987, two in 1988, two in 1989 and then one a year.

Orders may pick up during the late 1990s, when EDF will want to replace its first PWRs (which went on line in the mid-1970s), but meanwhile the company must either rely on the slack export market or diversify. Profits are expected to amount to FF 800 million (£7 million) over the next four years, but are somewhat unpredictable thereafter. Diversification into computer software and desalinization had begun before the takeover, but the size of the CGE group (second only to Renault in France) should now simplify redeployment of the 7,600 Framatome staff.

Framatome may also be better placed to compete for the few nuclear contracts now available abroad. Framatome has feared that competitors would be able to exploit — and were exploiting — its anomalous position after the Creusot-Loire collapse. But there must be doubt as to whether the now monolithic structure of the French nuclear enterprise will prove an asset or a hindrance.

In the past, governments and utilities have tended to diversify risks by placing orders for different parts of a power station (nuclear, conventional, civil) with different companies and even different countries, and it is pointed out in France that the almost equally monolithic German nuclear constructor, Kraftwerk Union, has not had an export order since 1975. In France, however, the now closer government ties with nuclear construction — although always close before — may not harm export prospects in what was always a highly politicized market.

Robert Walgate

Sakharov

Does no news mean bad news?

DR Andrei Sakharov may have been taken from his Gor'kii apartment to an unknown destination, a former Soviet dissident, the writer Lev Kopelev, claimed last week. Kopelev, who now lives in West Germany, told correspondents that the duty policeman who had been constantly stationed outside the Gor'kii apartment block in which Sakharov has been held had been removed in mid-August. When questioned by Dr Efrem Yankelevich, Sakharov's stepson-in-law and his representative abroad, Kopelev said that a recent visitor to Moscow had been told of it by a dissident source.

The last definite news of either Sakharov or his wife Elena Bonner was a postcard from the latter dated 4 July. This postcard, and one dated 29 June, broke an enforced silence of more than two months, during which, judging from the numbering of the postcards, 28 cards had been posted only to disappear *en route*. In both these cards, Mrs Bonner traced over the pronoun "Ya" (I) and the verbal endings indicating the first person singular, presumably to contrast, by implication, her condition ("I feel, physically, in good shape") with that of her husband, who at that time was on hunger strike in hospital.

At the end of June, the West German newspaper *Bild* came into possession of two videotapes showing Sakharov as a patient in the Semashko Hospital in Gor'kii, in which, *inter alia*, Sakharov was shown reading Western scientific articles and news magazines whose dates suggest that Sakharov was still alive in mid-June. A further video released through *Bild* indicated that Sakharov was discharged from the hospital on 11 July, but, says Dr Yankelevich, the evidence for the date was, in this case, only a notice on the hospital wall. According to the commentary, the *Bild* tapes were made specifically to refute Western fears that Sakharov was being medically mistreated.

Such reassurances were probably associated with the then impending tenth anniversary of the Helsinki Final Act on 1 August, and with the decision of the Soviet authorities, relayed to US Senator Paul Simon through the president of the Soviet Academy of Sciences, Dr Anatolii Aleksandrov, that it would be a violation of the Non-Proliferation Treaty to let Sakharov go abroad, since he could disclose atomic secrets to other governments and private organizations, "including terrorist groups".

If, indeed, the story of the vanishing policeman is true, and the Sakharovs have, indeed, been removed from Gor'kii, then the palliative was only temporary, and new fears for the couple's safety may well be justified.

Vera Rich

Pain and laboratory animals

SIR—The Mobilisation for Laboratory Animals campaign against the government's proposals to update the 1876 Cruelty to Animals Act as set out in the two White Papers (Commands 8883 and 9251) can, sadly, only delay the constructive reforms that are contained therein which undoubtedly would improve laboratory animal health and welfare. A further example of this is a recent letter you published¹.

While most, if not all, would prefer not to use animals for experimental and scientific purposes, unfortunately there are, at present, few valid alternatives and if advances in the treatment of human and animal diseases are to continue, animals will have to be used. In the meantime, efforts should be made by all those who use animals to refine experiments so that no unnecessary "suffering" occurs. (By suffering I include pain, distress, discomfort, stress and other such responses to the experimental conditions.) One must first, however, be able to recognize the signs of this "abnormal behaviour", an appraisal that practising veterinary surgeons and doctors carry out routinely. Furthermore, good stockmen and pet owners are also able to tell if their animals are unwell. While it may be criticized as being subjective, it is still of practical help to the animal, as appropriate treatment for relief can be instituted and monitored (the effectiveness of any treatment has also to be assessed).

It is extremely important for users of laboratory animals to be able to recognize normal and abnormal behaviour. The proposed liaison between the veterinary surgeon, the person with day to day responsibility and the user, as set out in the white papers, will do much to attain this goal. Pain is often an unwanted variable in the experimental situation and scientists would, therefore, be happy to eliminate it.

The recognition of pain and its assessment has been the subject of a recent publication² and a seminar to address this question was organized by the British Veterinary Association Animal Welfare Foundation (the proceedings will be published shortly). The topic of relief of pain in animals has also been addressed³. It has been disputed by Sharpe¹ that levels of pain cannot be recognized but I would take issue with this and would like to cite one of many examples of how this can be achieved. Lameness may present in animals in several ways; the animal may just show a limp (mild), the limp may have an evident effect on the movement (moderate), or the animal may hold the affected leg off the ground altogether (substantial). Thus the pain involved in this instance has been subdivided on clinical grounds and has obvious implications for the clinician. This can equally apply to

experimental animals and to other body systems and I would suggest that while one cannot be precise in the estimation of pain, or its quantitation, it is possible in many instances to predict and allocate pain into categories.

Where the government's recent proposals have so much to commend them is in their intention to control the amount of pain in experiments, as suggested in the BVA/CRAE/FRAME proposals, to balance pain against potential benefit. Three categories of pain (severity) have been proposed and for the first time a termination of experiment limit less than "severe and enduring pain" will have been instituted. There are many other such positive and constructive proposals in the white papers, for example control over breeding, supply and care of animals, a tighter licensing system, a strengthened inspectorate and a new statutory animal procedures committee. Regrettably a less positive proposal is the imposition of fees on an already overburdened and under-financed research budget.

In conclusion I would support the view that pain often can be categorized and the government's proposals should benefit both animal welfare and science in the long run.

D.B. MORTON

1. Sharpe *Nature* 316, 286 (1985).
2. Morton & Griffiths *Vet. Rec.* 116, 431-436 (1985).
3. Flecknell *Lab. Animals* 18, 147-160 (1983).

German universities

SIR—In Jürgen Neffe's article on West German universities (*Nature* 11 July, p.96) the main reason for students' protest against amendments to university regulations is mentioned explicitly. However, what your correspondent calls the "obviously reactionary character" of the suggested improvements should be explained in more detail.

The German federal *Hochschulrahmengesetz* (HRG) outlines a framework for regulations by the *Länder* only. Indeed, the *Land* of Hessen, in implementing the HRG of 1976, had shifted the balance within university committees to a wider group in 1978 and more drastically in 1980. One asserted aim of its regulations was to make it possible for professors to be outvoted, even when they are unanimous among themselves.

The combined representation of students, assistants and non-academic personnel was fixed at 55 out of a total of 90 delegates, leaving the remaining 35 for professors, who are thus a minority of 38.9 per cent. Consequently presidents and vice-presidents of Hessian universities have often been elected when 90 per cent of their professors have voted against them. Moreover, in elections for the delegates of the different groups, more than 85

per cent of professors, but fewer than 25 per cent of students and non-academic personnel, actually vote. The federal government intends to increase representation of professors to 50 per cent to make the framework comparable with that in other countries. Professors can still be overruled, but not if they are unanimous among themselves.

Government, *Bundestag* and *Bundesrat* are thus helping the universities to elect scientifically outstanding candidates as presidents and vice-presidents in Hessen and other German *Länder*. We do not consider this "obviously reactionary".

DIETER MEURER

WOLFGANG ZIMMERMANN

*Philipps-Universität Marburg,
Biegenstrasse 10,
D-3550 Marburg, FRG*

Spellbound

SIR—Your correspondent Monique Mulder (*Nature* 20 June, p.626), claiming some knowledge of the Kalenjin languages, suggests I misspelt the name of my colleague John Kimengich, as the spelling deviates from her bovine concept of its etymology. This is something over which we could argue until the cows come home, but in general I make it a rule to spell people's names in accordance with their own practice and not to worry unduly over conjectured derivation and meaning.

As your correspondent would have us believe that one of her names is "Borgerhoff", I commend this principle to her as one she would be well advised to encourage.

ANDREW HILL

*Yale University,
Department of Anthropology,
PO Box 2114,
Yale Station,
New Haven, Connecticut 06520, USA*

Chance encounters

SIR—We are collecting data on: (a) references in the literature to the role (however limited or extensive) of chance in the advance of science, engineering and technology; (b) personal experiences of serendipity unrecorded, or incompletely referred to in published papers.

We would be grateful for any well-documented information that readers can supply. Please send such material to the last-named.

PATRICK J. HANNAN

*Code 6180,
Naval Research Laboratory,
Washington, DC 20375-5000, USA*

JOHN F. CHRISTMAN

*Loyola University,
New Orleans, Louisiana 70118, USA*

RUSTUM ROY

*202 Materials Research Laboratory,
Pennsylvania State University,
University Park, Pennsylvania 16802,
USA*

The hunt for neutrino mass (cont'd)

New measurements deny an earlier report of a neutrino mass of 17.1 keV. But the hunt will go on. One benefit will be more accurate measurements of decay energy-spectra.

THE hunt for the heavy neutrino continues, with mounting ingenuity but without success. So much is clear from the latest experiment, reported by a group from the physics department at Princeton University (Altitzoglou, T. *et al.*, *Phys. Rev. Lett.* **55**, 799; 1985). The group's conclusion is that there is no evidence from the electron decay of sulphur-35 (^{35}S) for the presence of neutrinos with non-zero mass among the decay products. But given the compatibility of heavy neutrinos with present particle theories including the Salam-Weinberg electroweak theory, the tantalizing reports that heavy neutrinos may indeed be real and their attractiveness in other fields, cosmology in particular, it is unlikely that Altitzoglou *et al.* have spoken the last word on the subject.

For the time being, β -decay seems to have become the preferred phenomenon, if only because it allows human-scale experiments to be designed. The basis of these investigations remains substantially what it was in the 1930s, when Fermi postulated the existence of a neutrino with zero mass, but otherwise with the properties of electronic matter (leptons), to accommodate the observed energy spectra of the electrons (ranging from nothing to a maximum) so as to allow energy and linear and angular momentum (spin) to be conserved.

The difference now is that there are at least three generations of neutrinos, one corresponding to each of the three known leptons (muons and taus as well as electrons), and that even if the electron neutrino (strictly, it is an anti-neutrino) should have zero mass, the theories require that all three neutrinos (but there may be more) should be mixed together in the sense of quantum mechanics in the products observed.

The first experiment designed to look for non-zero mass was reported by a Soviet group nearly a decade ago, and turned on the measurement of the energy spectrum of the electrons from the β -decay of tritium. The result reported, a non-zero mass, has served chiefly as a reminder of the formidable difficulty of making convincing measurements in this field. The essential problem is that neutrinos with non-zero mass will affect the energy-spectrum of the electrons emitted in β -decay only in the region of electron-energy corresponding to the supposed mass, which cannot be more than a few electron-volts without being more of an

embarrassment than a comfort to the cosmologists. In short, the hunt for neutrinos with non-zero mass must be carried out just in that region where accurate measurements are most difficult.

Experimentalists are clearly undismayed by this challenge. Thus, earlier this year, J.J. Simpson, from the University of Guelph in Canada, described a repetition of the Soviet measurement by the intriguing technique of embedding radioactive tritium atoms in a slice of silicon, in which decay electrons can be counted as pulses of electric current, and where their total energy can be measured directly (*Phys. Rev. Lett.* **54**, 1891; 1985).

Simpson's conclusion was that the shape of his electron energy-spectrum was indeed distorted from that expected on the assumption that all neutrinos have zero mass; at the low end of the spectrum, there was an excess of electrons, most simply explained by a redistribution of the total energy shared by the products of the β -decay between the electron and the neutrino. His result was that at least one of the three neutrinos has non-zero mass, that the most probable value of that mass is 17.1 keV and that the quantum state representing the heavy neutrino constitutes some 3 per cent of that observed in experiments, manifested by the decay electrons.

As in true love, the path of discovery is never quite smooth. No doubt inspired by the uncomfortably large neutrino mass reported by Simpson, W.C. Haxton from the University of Washington and Los Alamos has already put forward one reason for believing that the result may be an artefact of interpretation (*Phys. Rev. Lett.* **55**, 807; 1985). Simpson's result hangs on the difference between the measured spectrum and that predicted in calculations with neutrinos of zero mass.

Haxton argues that Simpson's expectations may be oversimple approximations. His argument may be incomplete, but it is sufficient to show the complications with which people will have to grapple before unambiguous measurements can be made. Briefly, Simpson's argument seems to have been based on the standard approximations in the calculation of the rate of β -decay which, among other things, suppose that the electron that escapes from the decaying nucleus has no effect on the electron of the tritium atoms (which is transformed by the decay into a singly ionized helium atom). But electrons are in-

distinguishable on the scale of atoms, so may not the β -decay of tritium proceed by the capture of β -electron into a bound state and the transfer of the original bound electron into an essentially free state? Plainly the influence of these routes to β -decay will be greater when the electron-energy is small; Haxton argues that a first calculation shows them to be large enough to account for roughly one sixth of the departures from expectation reported by Simpson. Whatever the inevitable extrapolation of these arguments may bring, the result will be more accurate calculations of the detail of the decay of tritium than there have ever been before.

In the same spirit, whatever the immediate importance of the work of the Princeton group, the result will be a dramatic improvement in the accuracy of measurement of β -decay. Altitzoglou *et al.* have designed what is in effect a dedicated equipment for the direct measurement of β -decay spectra. The trick is to build a cylindrical vacuum chamber with a pair of symmetrically-placed coils to provide an axial magnetic field, and to arrange for the accurate positioning of a radioactive source and an electron-detector at the opposite ends of the axis. The chamber is iron-free, decay electrons are prevented from making the straight-line journey from source to detector by a lead obstacle in the centre, but can reach the detector if they happen to be orientated so as to intercept a circular annulus cut in an otherwise impenetrable disk, also concentrically mounted. The detector will then see only those electrons that leave the source with an energy or momentum exactly corresponding to the geometry of the equipment. The spectrum of a β -decay process is then determined by varying the current and recording the rate at which particles reach the detector.

The result is simply stated. For sulphur-35, the measured spectra fit better with the assumption that all neutrinos have zero-mass than with the particular result put forward by Simpson. On the face of things, neutrinos have no mass or, more precisely, such states of massive neutrinos as there may be play only a small part (or mix) in the states of the neutrinos emitted in β -decay. But this conclusion naturally applies only if neutrino masses are as large as suggested by Simpson; nothing so far attempted can exclude masses some orders of magnitude smaller than Simpson's 17.1 keV.

John Maddox

Virus structure

Crystallography attacks the cold

from Don C. Wiley

A LITTLE under eight years ago Harrison and co-workers published¹ the first three-dimensional structure of a virus after ten years of effort. Now it has taken Rossmann *et al.* the incredibly short period of thirteen months to determine the three dimensional structure of a rhinovirus — the causative agent of the common cold. Their achievement, reported on page 145 of this issue², is the result of a *tour de force* of modern X-ray crystallography, which is entering a new era. It is certainly nothing to sneeze at.

Perhaps the most intriguing discovery

to emerge from the structure is that the tertiary folds and quaternary organization of the polypeptides in rhinovirus are remarkably similar to those found in tomato bushy stunt virus³ and southern bean mosaic³ virus, both icosahedral plant viruses, indicating an evolutionary relationship between these infectious agents. It will be particularly interesting to compare the structure with that of poliovirus, which is soon to be published by Hogle and co-workers⁴. This will be the subject of a future contribution to these columns.

But will knowledge of the structure lead

to a cure for the common cold? The main reason that we do not develop immunity to colds, the way we do to childhood diseases like measles, and also the reason that no one has developed an effective vaccine against colds, seems to be that rhinoviruses exist in over eighty antigenically distinct serotypes (poliovirus, by contrast, has three). And each is different enough to escape antibodies against the other, but all are capable of delivering the same nasty cold. Through a collaboration with Rueckert's laboratory, Rossmann *et al.* describe four immunogenic sites on the virus that can generate neutralizing monoclonal antibodies in mice. The prospect of developing a 'comprehensive vaccine' against these sites seems remote, however, because the evidence indicates, presumably because of their exposed positions on external loops, that they can accept changes in structure (sequence) without consequent disruptions of any viral functions. So they may be able to stay ahead of the vaccine designers as well as our immune systems.

May there, instead, be a conserved epitope, shared by all rhinoviruses that could be a vaccine target? One prospect might have been the receptor-binding cavity, which binds the virus to a receptor on a target cell. Unfortunately, it seems to be in a deep cleft inaccessible to antibodies. A similar situation has been observed for the two glycoprotein surface antigens of influenza virus^{5,6}, where the conserved residues making up the receptor-binding pockets are either inaccessible to antibodies, or surrounded by a rim of amino acids that can vary and in so doing dislodge antibodies targeted at the conserved site. It remains to be seen whether rhinovirus carries some other common, conserved epitope that would not change structure under the selective pressure of antisera and against which neutralizing antibodies could be artificially stimulated (by a peptide?). In the meantime, as with the crystallographic studies of other infectious agents^{5,7}, basic issues such as the structural nature of immunogenic regions and the mechanism of immunological neutralization can now be studied in atomic detail.

Macromolecular crystallography is often misunderstood. It sometimes appears to be just a descriptive anatomy of macromolecules, providing complicated illustrations for textbooks. At its best, however, it provides the starting point for detailed research into how large molecules work. The three-dimensional patterns of chemical reactivity on a macromolecule are the key to chemical analysis of the capabilities of that molecule. For example, one can now foresee three non-immunological approaches to curing the common cold. First, knowing the structure of the receptor-binding site on the virus should stimulate efforts to design a tight-binding specific inhibitor of virus-cell attachment. Second, structural

Sir Frank MacFarlane Burnet (1899–1985)

MACFARLANE Burnet, joint winner of the 1960 Nobel Prize for Physiology and Medicine and originator of the clonal selection theory, which forms the basis of modern immunology, died peacefully on 31st August, 1985. In this technological era, we can ask whether anyone will ever sweep as broadly or dare as boldly as the man who dubbed himself "the last of the great amateurs". The essence of Burnet's creativity was his ability to move from the restricted, reductionistic framework of particular experiments to nature's grand design, always searching for the general biological principle.

Burnet's first interest was in bacteriophages and he was the first to realize that the hereditary phage particle integrates itself into the 'hereditary constitution' of the lysogenic bacteria as a non-infectious anlage which multiplies in step with the bacterial cell. Animal viruses, much harder to grow than phages, preoccupied Burnet for the next thirty years. He improved the techniques for growing viruses in chick embryo, established the chief quantitative method for studying viral replication and isolated the causative organism of Q fever — a rickettsia and not a virus.

Burnet's interest in immunology was a natural outgrowth of his broad approach to virology. Three observations intrigued him particularly. First, chick embryos could not be immunized, no matter how hard he tried. Secondly, mice infected *in utero* with lymphocytic choriomeningitis virus had been shown by Traub to carry virus in their blood and tissues for life, without forming antibody. Thirdly, Burnet was struck by Owen's observation that dizygotic twin cattle which shared a common placenta could be born with and continue to carry two types of red blood cell, their own and that of the twin, without an immune response. Burnet predicted that an antigen artificially introduced into an immature

immune system would lead to the animal being tricked into thinking that the antigen was not foreign. Billingham, Brent and Medawar provided the necessary experimental evidence to establish immunological tolerance as a reality and so Burnet and Medawar shared the Nobel prize for the discovery.

Burnet's most profound and unequivocally correct theoretical insight came some years after all the excitement about tolerance. Stimulated by Jerne's 'natural selection' theory, Burnet postulated the clonal selection theory in 1957. The essence of this is that lymphocyte cells bear on their surface specific immunoglobulin molecules reflecting the specificity of the antibody that will later be synthesized once the cell is activated. A process of "somatic mutation or conceivably other inheritable changes" leads to diversification such that each cell has only one specificity, but the population as a whole constitutes a repertoire "corresponding probably with varying degrees of precision to all, or virtually all, the antigenic determinants that occur in biological material". Antigen serves as a selective stimulus, causing preferential proliferation and differentiation of the clones that have receptors for that antigen. He saw self-tolerance as involving the deletion of clones with receptors for self antigens. The cellular and molecular vindication of this theory gave Burnet his most lasting satisfaction.

Burnet was a biologist. His love-hate affair with biochemistry, which led to a brief but damaging rejection of the worth of molecular biology (*Lancet* (i) 37; 1966), stemmed from his belief that biochemists should serve as superior servants to the biologists responsible for uncovering nature's real secrets. Despite this ambivalence, he sponsored some first-rate biochemical work. Above all he will be remembered as one who believed science is primarily about ideas. G. J. V. Nossal

studies of an 'uncoating transition' of the virus subunits, which may be involved in the process by which the virus delivers its genome across a membrane into the cytoplasm, could identify a target for a new therapeutic agent that would inhibit the entry and/or uncoating process. And third, an atomic description of the subunit interfaces and how they interact during viral assembly could similarly provide a target for an assembly-inhibiting drug. None of these obvious ideas is trivial to pursue, as those currently designing enzyme inhibitors will attest; but the atomic structure of rhinovirus opens them to a rational attack.

What makes this all realistic is a combination of rapid technological advances in macromolecular X-ray crystallography and molecular biology. The rhinovirus structure determination illustrates two major advances. First, all the X-ray data, over eight hundred thousand unique reflections, were collected using a synchrotron X-ray source — the Cornell High Energy Synchrotron Source. Synchrotron radiation is an extremely bright source of X rays. X-ray exposure times drop from many hours to minutes at a synchrotron; data collection times from months to days. Second, a powerful new method for determining the phases of high resolution X-ray data has been demonstrated. Normal procedures have required data from at least one isomorphous heavy atom derivative to high resolution (3 Å). Rossmann *et al.* were able to abandon heavy atom phases at low resolution (6 Å), and calculate high resolution phases by a stepwise (22 steps) extension of the low resolution phases using algorithms previously used for non-crystallographic symmetry-based phase

refinement. I think they are right to note: "The success of these techniques holds promise for a rapid series of structural determinations of viruses and other biological assemblies of high symmetries, by reducing dependence upon finding satisfactory isomorphous heavy atom derivatives."

The scientific and technological feat of the rhinovirus structure determination is only one in a series of recent events in macromolecular crystallography that have brought us many new structures: including the first of a membrane protein, and structures of protein-DNA complexes and an antibody-antigen complex. Advances in molecular biology make available large quantities of pure protein, and site-directed mutagenesis improves prospects for understanding structure-function relationship. An array of technical advances, ranging from computers, synchrotron sources, area detectors and computer graphics terminals to new algorithms for solving structures, have reduced the drudgery of macromolecular crystallography and have made it a 'high tech' operation whose critical input is a trained crystallographer and a good crystal. Despite this cheery outlook, it may be a while before we have heard the last sneeze. □

1. Harrison, S.C. *et al.* *Nature* **276**, 368 (1978).
2. Rossmann, M.G. *et al.* *Nature* **317**, 145 (1985).
3. Abad-Zapatero, Z. *et al.* *Nature* **286**, 33 (1980).
4. Hogle, J.M., Chow, M. & Filman, D.J. *Science* (in the press).
5. Wiley, D.C., Wilson, I.A. & Skehel, J.J. *Nature* **289**, 373 (1981).
6. Colman, P.M., Varghese, J.N. & Laver, W.G. *Nature* **303**, 41 (1983).
7. Freymann, D.M. *et al.* *Nature* **311**, 167 (1984).

Don C. Wiley is in the Department of Biochemistry and Molecular Biology, Harvard University, Massachusetts 02138, USA.

Catalyst chemistry

Silica sodalite without water

from Lovat Rees

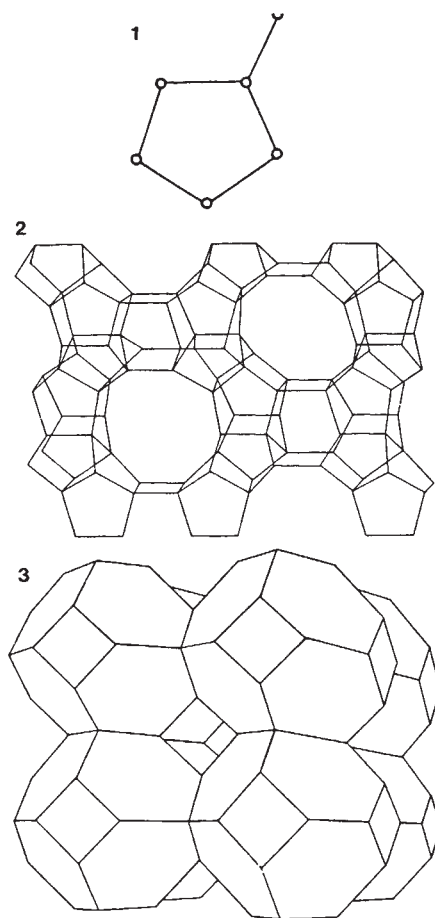
ZEOLITES are best known as the catalysts used in oil refineries for cracking crude oil. Their crystalline structure forms a three-dimensional network of molecule-sized channels, which also makes them ideal as molecular sieves. Starting from the pioneering studies by R.M. Barrer in the 1940s many aluminosilicate zeolites and, in some cases, their pure SiO₂ analogues have been routinely synthesized under hydrothermal conditions. On p157 of this issue, D.M. Bibby and M.P. Dale¹ report on an exciting new development in which the pure silica sodalite framework and its silica-rich aluminosilicate form is synthesized using ethylene glycol in place of water both as a solvent and a templating molecule. Although these compounds themselves have little commercial importance, they will contribute substantially to our understanding of the physical chemistry of zeolites.

Sodalite occurs widely in nature and is also readily synthesized hydrothermally with Al/Si ratios near unity. Silica-rich sodalite has been prepared from aqueous hydrothermal systems using tetramethylammonium templating cations. This system has also produced silica-rich zeolites such as ZSM-5 and ZSM-11 and their respective end-members, silicalite-1 and silicalite-2, that contain essentially no aluminium²⁻⁴. These are rich in the 5-membered rings that make up the silica-rich framework. For example, the only secondary unit to build the ZSM-5 and silicalite-1 structure is the 5-1 unit shown in Fig. 1 which can be assembled into the three-dimensional structure of ZSM-5 of Fig. 2.

The silica-sodalite synthesized by Bibby and Dale is unusual because the framework contains only 4- and 6-membered rings, as can be seen in Fig. 3.

Such structures are characteristic of high-alumina content framework silicates synthesized from aqueous systems using inorganic bases. The most important zeolite of this type is zeolite-Y which is the cracking catalyst used in most oil refineries. Much of the technology behind the use of this catalyst has been concerned with the inability to synthesize this framework with Si/Al ratios in excess of 3, and catalyst manufacturers have put considerable effort into reducing the Al content to enhance its thermal stability while retaining catalytic activity. If the direct synthesis of a silica-rich form of zeolite-Y could be achieved using an organic solvent or an organic solvent plus a large organic templating ion, the problems associated with the production of dealuminated zeolite-Y could be eliminated, although the process might turn out to be costly.

Silica-sodalite and its silica-rich aluminosilicate forms are not particularly useful as catalysts, as the 'windows' controlling access to the channel network of the sodalite structure are too small. With a free diameter of ~0.21 nm they allow only rapid



ingress of water molecules and some small, not too highly hydrated, cations. But these materials have great importance from the point of view of fundamental physical chemistry. First, silicalite-1 has been shown to be hydrophobic in nature. The sorption of water is very limited compared with its sorption capacity for organic molecules, such as hydrocarbons.

Comparison of the sorption of water in silica-sodalite with its different pore-size characteristics with that in silicalite-1, which has two sets of channels, may lead to a better understanding of the reasons for the lack of hydrogen-bonding that must occur between water molecules sorbed in the silicalite-1 framework.

Second, since Engelhardt *et al.*⁵ introduced solid-state magic-angle spinning NMR (MASNMR) six years ago, our knowledge of zeolite frameworks and the distribution of Al in these frameworks has been greatly enhanced. It is now apparent, however, that the complete analysis of these NMR spectra for structures that are not cubic and that contain more than one type of tetrahedral Si/Al site can only be accomplished if the spectrum of the pure silica form of the framework is available. Some success in producing pure silica frameworks has been achieved by dealumination of the aluminosilicate zeolites, but the process involves the annealing of many vacancies in the framework produced by the removal of Al atoms; it leaves material containing meso-

pores that must result in considerable lattice strain energy. The direct synthesis of some of these pure silica frameworks using organic solvents would enable a considerable advance in our use of MASNMR.

The spectra which have been published to date suggest that a simple relationship between Si-O-Si bond angles and chemical shifts must exist⁶. We await the production of good pure-silica framework crystals for the complementary X-ray and MASNMR studies required to establish this relationship. Such basic knowledge would be of immense significance in advancing our understanding of the chemistry of zeolites and other aluminosilicates. □

1. Bibby, D.M. & Dale, M.P. *Nature* **317**, 157 (1985).
2. Flanagan, E.M. *et al.* *Nature* **271**, 512 (1978).
3. Kokotalia, G.T., Chu, P., Lawton, S.L. & Meier, W.M. *Nature* **275**, 119 (1978).
4. Bibby, D.M., Milestone, N.B. & Aldridge, L.P. *Nature* **280**, 664 (1979).
5. Engelhardt, G. *et al.* in *Workshop on Absorption of Hydrocarbons in Zeolites* (Acad. Sci. GDR, Berlin, 1979).
6. Smith, J.V. & Blackwell, C.S. *Nature* **303**, 223 (1983).

Lovat Rees is Reader in Physical Chemistry, Imperial College, London SW7 2AY, UK.

Particle physics

The search for truth at Bari

from David J. Miller

TWO KINDS of 'truth' provided the major questions at the Bari High-Energy Physics conference*. Most interest was generated by the new 'Theory of Everything' — the superstring theory. It is a profound and beautiful piece of mathematics, but is it true? Observation of the 'truth' or 'top' quantum number is still not confirmed by experimenters at the CERN proton-antiproton collider. There were also useful warnings about ideas for future accelerators, three counter-results to a recent measurement of a large mass for electron-neutrinos, scepticism about the neutrinos which may be coming from Cyg X3, and a lot of sound data that reinforces the standard model of strong (quantum chromodynamics) and electroweak interactions with three generations of quarks, charged leptons and neutrinos.

A number of results reported from the CERN collider by the UA1 and UA2 experimental collaborators last year (see *Nature News and Views* **311**, 210; 1984) have not been confirmed by the new data. The identification of top or truth is particularly important because it is the only quark missing from the three generations. The UA1 experimenters do indeed have more events in the kinematic region where they claimed to find the signal for a W^+ boson decaying to top and antibottom quarks. In fact, they have too many events. There might be additional direct

production of top-antitop pairs but if so, why do all of the events cluster close to the W^+ mass? L. DiLella (CERN), a member of UA2, reminded us that they still have not seen any convincing top-quark signal, and suggested that the current upgrading of both the detectors will help them sort out what is happening. A more secure, and very important, result is a measurement of the width of the Z^0 boson. Combining UA1 and UA2 results, it has been possible to use this width to put a limit of 2.4 on the number of extra generations of light neutrino, beyond the three of the standard model, with 90 per cent confidence (experimental) and with a theoretical uncertainty of ± 1 .

It was difficult to get the particle theorists to talk about anything else but superstring theory. There is still a lot of work to be done according to M. Green (Queen Mary College, London), one of its inventors, but it has already been successful in overcoming the problem of anomalies at the Planck scale, where all previous attempts at a quantum theory of gravity have failed. Its fundamental symmetry is that of vibrating strings in a ten-dimensional space. Six of these dimensions have to be curled up to regain four-dimensional world, but there should then be no free parameters. The arbitrary features of the standard model, such as the number of generations, fermion and boson masses, should all be determined by the properties of the Calabi-Yao manifolds in the six curled-up dimensions. Unfortunately,

there are ten thousand Calabi-Yao manifolds, and it is taking a little while to pick the one that corresponds to the properties of the particles we know. But when it is picked, as one theoretician pointed out, experimental particle physicists may all be able to retire.

Experimenters, used to the rapid appearance and disappearance of ambitious theories, were more concerned with immediately testable predictions. So far, the only predictions are the same as those of supersymmetry, which is contained within superstring theory, but despite searches for supersymmetric effects nothing definite has been seen.

The sure way to find out what lies beyond the standard model is to raise the experimental energy into the region where the Higgs fields, or their supersymmetric equivalents, must become accessible. A number of speakers, notably J. Lawson (Rutherford Appleton Laboratory), demonstrated that hard and calculable limits exist to what can be achieved by any accelerating device. To use the high electric fields that might be available from a 'wake-field' or a 'beat-wave' accelerator, it will be necessary to develop new beam-focusing techniques, more efficient lasers or radio-frequency cavities and much brighter particle sources than are now available. What is needed is a fundamental development — like the discoveries of strong-focusing or stochastic cooling. Fundamental problems in a colliding-beam machine include 'beamstrahlung', the energy loss of particles in one bunch due to the collective magnetic field of the opposed bunch. There is a possibility of overcoming this for the majority of particles in each bunch by using quantum effects, but only if the beam bunches are less than a micrometer long and a few angstroms across. No one yet knows how to make such tiny beams.

The Kendrew report on high-energy particle physics in the United Kingdom was debated briefly. The suggestion that particle physicists should collaborate more with industry and with other sciences was well received, and in the accelerator and detector sessions many examples of such collaboration were reported—such as the phototriodes developed by the Rutherford Appleton Laboratory in collaboration with Philips. Kendrew's negative conclusions about the rate of expenditure on the field were thought not to have had too much influence outside Britain, although in his final review talk A. deRujula (CERN), commenting on the apparent rapid reduction in flux of the pulsed gamma-ray source in Cyg X3, suggested that the Galactobritannic Empire had decided to close down the accelerator, just as we thought we might be detecting its neutrino beam. □

David J. Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

* International Europhysics Conference on High-Energy Physics organized by the European Physical Society, Bari, Italy, 18–24 July 1985.

Conservation biology.

A discipline with a time limit

from Jared M. Diamond and Robert M. May

CONSERVATION biologists have long been seeking to apply basic research techniques to environmental problems. Their progress in this direction was the subject of a recent meeting*, soon to be published¹.

To show the scale of problems facing conservation biologists, Norman Myers reviewed the status of primary tropical rain forests. This type of habitat accounts for only 7 per cent of the Earth's surface but at least 50 per cent (possibly much higher) of its species and about 70–90 per cent of present-day extinctions. The remaining tropical rain forest, about 9 million km², is being destroyed at an accelerating rate that may exceed 200,000 km² (an area exceeding that of the United Kingdom) each year.

This destruction is caused mainly by small-scale subsistence agriculture, commercial logging and large-scale commercial ranching. During the next five years, three isolated rain forests with high human population density, a large number of endemic species and over 100,000 species altogether will probably be destroyed completely: the rainforests of Madagascar, Ecuador's Pacific coast and Brazil's Atlantic coast. Within the next 10–15 years distinctive rainforests in Central America, Southeast Asia, West Africa, the Himalayan foothills and Pacific islands will have largely disappeared, accounting for another 1 million km² and half a million species lost. In sixty years' time, only the Congo Basin and Western Amazonia will support extensive rain forests. Alwyn Genry (Missouri Botanical Gardens, St Louis) described how the recent biological exploration of Ecuador's isolated Centinela ridge revealed up to 90 new plant species confined to that ridge. But the ridge today has had its forest completely cut, and the species are gone. Thus, the present generation may well be the last to live in a species-rich world.

To reduce this extinction wave requires not only political clout but also biological understanding. Governments will provide some support for conservation, but not enough to save everything. How can this support best be used? The conference illustrated a wide spectrum of approaches, from detailed single-species studies to broad biogeographical approaches.

At one extreme, detailed discussion of the effects of inbreeding — which must always be a worry when dealing with small populations in zoos or reserves — took place against the background of a review by Kathy Ralls (National Zoo, Washington DC), Paul Harvey and Anna Marie Lyles (Princeton University) of evidence

that inbreeding does indeed depress reproductive success. Alan Templeton (Washington University, St Louis) described intensive studies that focus on the extent to which the amount of genetic variability within a particular target population or species may be manipulated. For example, all the extant Speke's gazelles in zoos derive from one male and three females; the current concern is to minimize the degree of inbreeding by careful design of the mating programme.

Another study dealt with reintroduction of a particular lizard species into scattered and isolated glades in the Ozark Mountains in the United States. Surviving populations of this lizard are homozygous and distinct in each glade, and a colonizing population made up by taking one or two lizards from each of several other glades tends to exhibit outbreeding depression in the short run, as the relatively successful homozygous gene complexes are reshuffled. Arguing from laboratory studies on *Drosophila*, Templeton suggested that, provided the introduced populations can be maintained for the first few generations of rearrangement of the gene pool, the long-term prospects were likely to be good, with the eventually heterozygous populations of lizards having higher average fitness than the homogeneous isolates from which they were constituted. In the very long run, of course, these new isolated populations would probably lose their heterozygosity, with genetic drift in small populations leading to the kind of homozygosity found in the currently existing populations. In summary, although inbreeding is a problem for many conservation programmes, careful design (often using information gathered by modern molecular techniques) may often be able to solve these problems.

At the opposite extreme, the critical biological information required to design the national park system for the Irian Jaya province of Indonesia (West New Guinea) has been gross and biogeographical: what areas harbour distinct sets of species? Given the government's intention to divide the province into conservation areas, timber or mining leases and population resettlement areas, biologists had to give advice immediately, based on whatever knowledge was already available. Detailed information about habits of thousands of species in each conservation area was superfluous. Indeed, the Kumawa Mountains were recommended as a reserve before any biologist had even been there, or before the occurrence of any particular species there had been recorded. The decision was instead based simply on their location and the consequent expect-

tation (later proved correct) that they harboured endemic species.

Habitat fragmentation is the result of destructive processes which reduce total area and split the remaining habitat into isolated small pieces. This constitutes the main threat to species in the temperate zones today. Its importance has also been demonstrated by the World Wildlife Fund's minimum area project in Amazonian tropical rain forest, where extinctions can be documented within a year of fragmentation². David Wilcove (Princeton University) and Thomas Lovejoy (World Wildlife Fund) discussed at least five proximate mechanisms that can cause such extinction:

(1) Species requiring big territories, such as large predators, disappear from fragments smaller than one territory.

(2) Loss of habitat heterogeneity in small fragments eliminates habitat specialists and species dependent on several habitats. According to the estimates of Steven Hubbell (University of Iowa), one-quarter of the tree species of Panama's Barro Colorado reserve are 'subsidized' by surrounding similar forests and would disappear (by virtue of these first two mechanisms) if that surrounding forest were cut down.

(3) Species found in surrounding but contrasting habitats, such as predators and nest parasites, may harm species near the edge of the enclosed habitat — the smaller the habitat area, the larger the edge-to-area ratio, and the more significant these effects become. For instance, by placing artificial nests with fresh quail eggs in woodlots of various sizes, Wilcove documented up to 95 per cent predation on nests in the smallest lots as opposed to only 2 per cent in the largest lots³.

(4) Other edge effects such as changes in humidity and temperature cause increased tree mortality, leaf-fall and other changes in forest fragments.

(5) Extinctions from any of these four factors cause secondary extinctions, such as the chain of events leading from the arrival of myxomatosis in Britain to the extinction of the Large Blue butterfly⁴.

John Terborgh (Princeton University) has studied mammalian frugivores in the Peruvian rain forest and illustrated how basic biological studies may help to promote compromises between economic development and conservation. Fruit production in the Manu National Park exhibits 20-fold seasonal variation. For 9 months, more fruit grows than can be consumed by the community of frugivorous mammals and birds (40 per cent of which, by weight, are primates in this undisturbed region). The shorter dry season is a period of fruit scarcity, however, during which 80 per cent of the animal biomass is sustained by fruit from only 10 'keystone plant species', comprising less than 1 per cent of the plant species in the area. These crucial resources are palm nuts, figs, and nectar. The usual forestry

*The 1985 symposium on conservation biology was held at the University of Michigan, USA from May 5–8 1985.

practice in the tropics is to clear-cut rain forest and to plant exotic commercial timber species that do not support fruit-eating mammals. Instead, as a few plant species control mammal population densities, Terborgh suggests mixed silviculture: harvest some of the trees that produce excess fruit in the wet season, replace them with commercial timber species and maintain the keystone tree species.

Perhaps the most important conclusion of the conference had to do with the changing needs of population biology and conservation biology since the last conference in 1978 (ref. 5). Population biology (including ecology, evolutionary biology, systematics and biogeography) is a basic science, which stands in the same relation to the applied sciences of restoration ecology and conservation biology as molecular and cellular biology stands to medicine⁶. There is a big communications gap between basic population biologists and land managers, and little organized or financial support to close this gap. Government agencies responsible for land management are forced to spend their limited funds on crisis control, rather than on closing this gap or on basic studies.

Although little time is available to close the gap — many of the species that we want to conserve will be extinct within five years — and conservation biology is the

sole field of biological study with such a time limit, the National Science Foundation (NSF) has no panel, research grants or traineeships in conservation biology; universities have no departments and few courses in the subject. The basic science of population biology on which conservation biology depends is starved for support — the NSF panel on ecology and systematics has an annual grants budget of about 3 million dollars, compared with the National Institutes of Health annual budget of over 4,000 million dollars for medical research. It is as if we were to approach human health problems without medical schools, with minute support for basic research in molecular and cellular biology and without organizations to foster communication between the few remaining molecular biologists and physicians. □

1. Soule, M.E. *Conservation Biology: Science of Diversity* (Sinauer, Sunderland, in the press).
2. Lovejoy, T.E. et al. in *Extinctions* (ed. Nitecki, M.H.) (University of Chicago Press, 1984).
3. Wilcove, D.S. *Ecology* (in the press).
4. Ratchliffe, D. *New Scientist* 457, 8 November 1979.
5. Soule, M.E. & Wilcox, B.A. *Conservation Biology: An Evolutionary-Ecological Approach* (Sinauer, 1980).
6. Diamond, J.M. *Nature* 313, 629 (1984).

Jared M. Diamond is Professor of Physiology, University of California, Los Angeles, California 90024, USA; Robert M. May is class of 1877 Professor of Zoology, Princeton University, New Jersey 08544, USA.

Material sciences

Searching for supermagnets

from Robert W. Cahn

LODESTONES have long been superseded by increasingly esoteric compounds in the search for magnetic power. The latest novelty is a class of Fe-Nd-B compounds which form the most powerful permanent magnets yet created. The performance of these is based on the tetragonal material Nd₂Fe₁₄B which, since its identification in late 1983, has been the subject of over a hundred papers. There has never before been a comparable worldwide burst of effort to produce, understand and improve a new magnetic material.

As described by Robinson¹, the starting point was work done in 1973 by Arthur Clark of the US Navy. The issue was how to improve on the then best available magnets, Sm₂Co and Sm₂Co₁₇. The basis of such magnets is that the rare-earth metal provides high crystal anisotropy and therefore high coercivity, while the transition metal raises the Curie temperature high enough to avoid restrictions on use. Attempts to replace cobalt with iron in the hope of enhancing the magnetization without sacrificing coercivity were generally unimpressive, but RFe₂ (where R = Tb, Dy or Sm) compounds were much better, provided they were prepared by making and then crystallizing a glassy structure. Evaporation was in due course

replaced by melt-spinning a jet of molten metal against a spinning copper wheel, using boron to stabilize a glassy structure. By the beginning of 1983 Fe-Tb-La-B and Fe-Nb-B ribbons had been prepared exhibiting high coercivity — the source of which was a mystery. Then it was discovered that the X-ray diffraction patterns from melt-spun and annealed Fe-Pr-B-Si ribbon showed a previously unknown phase, identical with one which had been observed, but not analysed, in 1979 in the Fe-Nd-B system; within a few months this phase was identified as Nd₂Fe₁₄B and its structure determined^{2,4}.

In June 1983, the Japanese company Sumitomo announced the production of Fe-Nd-B permanent magnets by the straightforward sintering route which works so well for the Co-Sm magnets, though for the new material the heat-treatment after sintering is critical. The compositions of both the sintered and the melt-spun versions of Fe-Nd-B have to be displaced from Nd₂Fe₁₄B stoichiometry for best magnetic properties, though the displacement is different for the two production routes. To summarize a great deal of research published since mid-1983, the compositional displacements seem to be needed to permit other, ultrafine, phases,

both magnetic (α -Fe) and non-magnetic (Fe₃NdB₂ and a Nd-rich phase) to precipitate at the boundaries of the Nd₂Fe₁₄B grains. This adds the effects of domain-wall pinning to the large magnetocrystalline anisotropy of the main phase.

For sintered magnets, domain-wall pinning is certainly crucial, since the grain size (5–15 μ m) is much larger than the 0.3 μ m diameter calculated as the 'single-domain particle diameter', D_c . However, for the melt-spun material (which, it seems, always passes through the glassy phase at some point in the process) the grain size, controlled by the quench rate, that gives optimum properties is 20–80 nm, considerably smaller than D_c . But the recognized experts on the rapid-quenching route at General Motors are convinced that the optimum grain size is in some way connected with the achievement of single-domain morphology, in spite of the mismatch with D_c . The relation between microstructure and magnetic properties was thoroughly analysed in a very up-to-date review by J.D. Livingston⁵ presented at a recent meeting on rare-earth magnets*, and also in a paper by Hadjipanayis and colleagues⁶, first presented at the November 1984 '3M' conference, at which the Japanese also gave an account of their sintered Fe-Nd-B material⁷.

Since their original account* in early 1984, the General Motors team have said little, but very recently Lee has published an account of the last year's developments⁸. One of the weaknesses of the melt-spun material (whether dispersed in a non-magnetic matrix or hot-pressed) has been its magnetic isotropy. Now, by die-upsetting (a process in which transverse plastic deformation is allowed) following hot-pressing, strongly anisotropic material can be prepared, with a much enhanced level of residual magnetization. The fact that Nd₂Fe₁₄B turns out to be plastic at 700°C is one of the many surprising features of this compound.

European physicists and metallurgists

*8th International Workshop on Rare-Earth Magnets and their Applications, incorporating the 4th International Symposium on Magnetic Anisotropy and Coercivity in Rare-Earth-Transition Metal Alloys, University of Dayton, Ohio, 6–9 May 1985.

1. Robinson, A.L. *Science* 233, 920 (1984).
2. Herbst, J.F., Croat, J.J., Pinkerton, F.E. & Yelon, W.B. *Phys. Rev. B* 29, 4176 (1984); *J. appl. Phys.* 57, 4086 (1985).
3. Givord, D., Li, H.S. & Moreau, J.M. *Solid State Commun.* 50, 497 (1984).
4. Shoemaker, C.B., Shoemaker, D.P. & Fruchart, R. *Acta Cryst.* C40, 1665 (1984).
5. Livingston, J.D. in *Proc. 8th Int. Workshop on Rare-Earth Magnets and their Applications*, 423 (Univ. Dayton, 1985).
6. Hadjipanayis, G.C., Lawless, K.P. & Dickerson, R.C. *J. appl. Phys.* 57, 4097 (1985).
7. Sagawa, M., Fujimura, S., Yamamoto, H., Matsuura, Y. & Hirotsawa, S. *J. appl. Phys.* 57, 4094 (1985).
8. Croat, J.J., Herbst, J.F., Lee, R.W. & Pinkerton, F.E. *Appl. Phys. Lett.* 44, 148 (1984).
9. Lee, R.W. *Appl. Phys. Lett.* 46, 790 (1985).
10. Coey, J.M.D., Cadogan, J.M. & Ryan, D.H. in *Proc. Nd-Fe Permanent Magnets — Their Present and Future Applications* (ed. Mitchell, I.V.) 143 (Commission of the European Communities, Brussels, 1984).
11. Fruchart, S. et al. in *Proc. Nd-Fe Permanent Magnets — Their Present and Future Applications* (ed. Mitchell, I.V.) 173 (Commission of the European Communities, Brussels, 1984).

have not been left out of this new field, on the evidence of the proceedings, only recently available, of a one-day 'workshop meeting' held in Brussels on 25 October 1984[†]. The influence of economics and the supply position is particularly clearly set out in this report — the relative scarcity of both cobalt and samarium stimulated the search for alternatives to Co-Sm magnets — and it is clear that neodymium, the most plentiful of the 'rare' earths, is not rare at all, being more abundant than lead. Perversely, the most serious practical difficulty with Nd₂Fe₁₄B — the low Curie temperature, T_c — can be readily overcome by adding cobalt. Of course, not only does this imply a renewed dependence on scarce cobalt supplies, but it also lowers the saturation magnet-

ization. An intriguing alternative^{10,11} that may have industrial potential is to raise T_c and magnetization by absorbing hydrogen into Nd₂Fe₁₄B; Fruchart *et al.*¹¹ report a T_c increase of 87 K.

Several papers presented at Brussels or Dayton dealt with possible applications of Fe-Nd-B based magnets, both within and outside the motor industry. The size of the research programme mounted by General Motors is sufficient indication of the potential for the use of high-grade permanent magnets, and these remarkable new materials offer scope for both fundamental research into the causes of magnetism and for profitable development. □

Robert W. Cahn, formerly Professor of Materials Science at the University of Sussex, is currently Visiting Research Fellow at the General Electric R & D Center, Schenectady, New York 12301, USA.

[†]"Nd-Fe Permanent Magnets — Their Present and Future Applications", 25 October 1984. Held at the Commission of the European Communities, Brussels. Report obtainable from Dr. I.V. Mitchell, CEC, Brussels.

Muscle contraction

Quick-frozen crossbridges

from Maxine Clarke

ON PAGE 182 of this issue, S. Tsukita and M. Yano publish the first clear electron micrographs of crossbridges in contracting muscle¹. This result has been sought ever since the importance of crossbridges in force generation was first realized over 20 years ago.

In the most widely accepted model of muscle contraction, crossbridges (projections of myosin heads from the thick filaments of muscle fibres) attach to the interdigitating thin filaments (actin) at right angles, then swing round to make an acute angle with the thick filament, causing the filaments to slide past each other and producing force. This basic mechanism and its modification as a result of new evidence was discussed recently in these pages².

Early work using insect flight muscle produced clear electron micrographs of crossbridge structure in relaxed and rigor (ATP-depleted) fibres. These pictures show crossbridges orientated at angles of about 90° (relaxed) and 50° (rigor) to the thick filament backbone³; the two orientations are thought to be equilibrium states of resting and maximal force-generating crossbridges.

Attempts to observe crossbridges of contracting muscle in the electron microscope have been much more problematic. No clear evidence of their structure has been observed until now, probably because fixation procedures cause gross disorder of the crossbridge arrangement, and preparative methods dehydrate the specimen, so that physiological conditions are not preserved in the micrographs. More recent efforts to elucidate crossbridge structure have therefore concentrated on characterization of other equilibrium states by the introduction of ATP ana-

logues, or on the study of isolated thick filaments. Recent progress in the latter has allowed crossbridges to be discerned in three-dimensional reconstructions of thick filaments⁴ (see figure).

Set against this background, the technical achievement of Tsukita and Yano is superb. They have produced electron micrographs of contracting mammalian skeletal muscle fibres which show that the crossbridge arrangement is ordered and similar to that in rigor muscle¹. The angle of crossbridge attachment may be slightly nearer to 90° than in rigor, but the data are

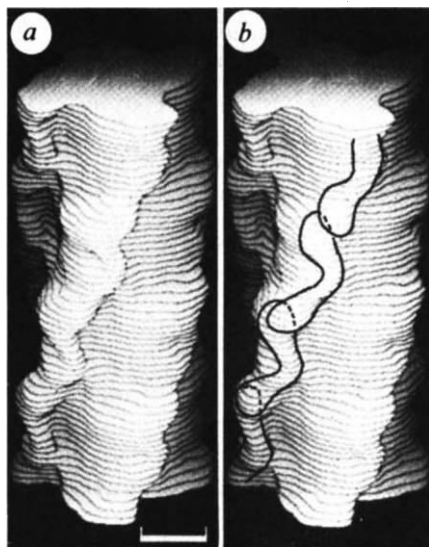
not yet accurate enough to characterize the configuration exactly.

What accounts for the success of Tsukita and Yano is, in part, that they have rapidly frozen actively contracting muscle by dropping the specimens on a copper block immersed in liquid helium, a technique developed by Heuser⁵. This process takes less than 2 ms, so there is an excellent chance that the crossbridges are preserved in physiological conditions. Unlike conventional electron microscope preparative techniques, rapid freezing ensures that the specimen is not dehydrated. The longitudinal section of muscle examined by Tsukita and Yano are only 15 nm thick, enabling a single layer of thick and thin filaments to be visualized.

More detailed information about crossbridge behaviour can be obtained by analysing optical diffraction patterns from slightly thicker sections of muscle fibres. Diffraction patterns of muscle fibres show a characteristic layer-line structure which arises from the helical periodicities of the thick (myosin) and thin (actin) filaments. In relaxed muscle, the myosin-based pattern is strong whereas in rigor the actin-based pattern dominates. The 14.3 nm reflection, strong in relaxed muscle, measures the crossbridge repeat with respect to their thick filament origin; the 37 nm reflection, strong in rigor muscle, measures the crossover repeat of the actin filament double helix. In X-ray diffraction patterns of contracting vertebrate muscle, most of the actin-based pattern disappears, which is not what the crossbridge theory predicts, but recently an enhancement of the intensity of the 5.9 nm actin reflection has been discerned^{6,7}. This is consistent with at least some ordered attachment during contraction, as demanded by the theory, but reflections from other actin-based layer lines still only provide support for the attachment of 20 per cent of the crossbridges, at most⁸. Equatorial reflections, which measure the transverse spacing between the filaments, clearly show that the bulk of the mass of the crossbridges is near the thin filaments in contracting muscle.

Optical diffraction patterns obtained by Tsukita and Yano from contracting muscle clearly show the 14.3 and 37 nm reflections. The authors interpret the strong 37 nm reflection to mean that most of the crossbridges are arrayed along the actin filaments, but the enhanced 14.3 nm reflection relative to rigor shows that the two states are not identical, contracting muscles having a larger thick filament-related component. The 37 nm reflection has not been reported in X-ray diffraction patterns. A reason for this could be that in frog muscle there is some 'smearing' of the myosin-based 42.9 nm reflection which obscures an enhanced 37 nm intensity.

Tsukita and Yano's highly ordered pictures of contracting crossbridges are consistent with the mechanical properties of contracting frog muscle. Stiffness, a para-



Surface representation of a tarantula muscle thick filament. In *b* the suggested arrangement of the myosin heads in the filament is superimposed. The distance between the contour levels is 1 nm; scale bar, 10 nm. From ref. 4, with kind permission of R. Craig.

meter that measures the number of attached crossbridges, is high both in rigor and in contracting muscle⁹.

Now that force-generating crossbridges have been seen, Tsukita and Yano can investigate further whether the structure of these crossbridges is similar to rigor or whether a distinct configuration can be characterized as the measurement technique is improved. One way to examine active crossbridges is to use techniques developed by Reedy and colleagues¹⁰ to obtain a three-dimensional picture of the crossbridge by tilting the electron micrographs and analysing the differently angled views of the section. In addition, the structure of the crossbridges can be compared with their predicted orientation

calculated from X-ray diffraction patterns. These techniques should soon allow a detailed view of the molecules responsible for producing force. □

1. Tsukita, T. & Yano, M. *Nature* **317**, 182 (1985).
2. Irving, M. *Nature News & Views* **316**, 292 (1985).
3. Reedy, M.K., Holmes, K.C. & Tregear, R.T. *Nature* **207**, 1276 (1965).
4. Crowther, R.A., Padron, R. & Craig, R. *J. molec. Biol.* **184**, 429 (1985).
5. Heuser, J.E. *J. molec. Biol.* **169**, 123 (1983).
6. Matsubara, I. *et al. Nature* **312**, 471 (1984).
7. Huxley, H.E., Kress, M. & Simmons, R.M. *Biophys. J.* (in the press).
8. Huxley, H.E. & Kress, M. *J. Mus. Res. Cell Motil.* **6**, 153 (1985).
9. Goldman, Y.E. & Simmons, R.M. *J. Physiol., Lond.* **269**, 558 (1977).
10. Taylor, K.A., Reedy, M.C., Cordova, L. & Reedy, M.K. *Nature* **310**, 285 (1984).

Maxine Clarke is an assistant editor of *Nature*.

Ionospheric physics

How to make a long antenna

from M.J. Rycroft

MEN have long felt the need to communicate with others around the world. New types of transmitting antennas are now becoming available for global communications, some of which are so unconventional that electrical engineers might not even recognize them as such. Ideally, the antenna has to be of the same order of size as the wavelength of the radio signal being transmitted. For signals in the extremely low frequency band (ELF, 3–3,000 Hz), which is used for communicating around the world and with submarines, this means that the antenna has to be extremely large, between 100,000 and 100 km. One solution is to use the Earth's atmosphere itself as an antenna. On page 155 of this issue, Barr *et al.* present the first quantitative results from a system that causes the ionosphere to act as an ELF transmitter by modulating the conductivity of the polar electrojet.

The lowest frequencies are chosen for radio signals when the data rate is low and when communication is required over long distances or under the sea. At ELF, the skin depth is largest, enabling radio waves to be received on submarines (D. Llanwyn Jones, *New Scient.* 4 July 1985, 37). A transmitter in Wisconsin uses a system of frequency modulation centred on 76 Hz for world-wide communication with submarines. Although the efficiency of such transmitters is low, $\sim 2 \times 10^{-6}$, the low attenuation rate of signals propagating in the Earth-ionosphere waveguide, about 1.5 dB/Mm (thousand km) at this frequency, enables even a relatively weak signal to be received anywhere on the surface of the Earth.

Wires long enough to act as transmitters for ELF signals do exist. In industrialized countries, power stations generating hundreds of MW serve an electrical grid system which covers an enormous geographical area. Such conductors, which

are typically 1 Mm long, are effective radiators of the fundamental AC (50 or 60 Hz) power signal. Its higher frequency harmonics extend into the VLF band (3–30 kHz), and propagate through the ionosphere. They are guided into the opposite hemisphere along geomagnetic field-aligned ducts of enhanced plasma density, which can be received as whistlers; these provide useful diagnostics of the plasma distribution in the magnetosphere.

Nature provides, *gratis*, power sources that are even larger than man can produce. The auroral electrojet current is a $\sim 10^6$ MW system driven by the solar wind flowing past the geomagnetic field via geomagnetically field-aligned currents (Birkeland currents). This essentially DC current flows in the ionosphere at a height near 100 km around the auroral oval. The auroral oval is a ring around the geomagnetic pole, normally near 67° geomagnetic latitude, where the ionospheric conductivity is enhanced by the precipitation of keV electrons that produce secondary ionization. These electrons also excite the molecules and atoms of the tenuous uppermost atmosphere to produce their characteristic optical radiation — the aurora. The atmosphere at these heights is heated, typically by 100 K, by ohmic dissipation of the ionospheric auroral electrojet currents.

The theoretical (Gurevich, A.V. *Non-linear Phenomena in the Ionosphere*, Springer, Berlin; 1978) and experimental (for example, Cannon, P.S. *J. atmos. terr. Phys.* **44**, 819; 1982) investigation of the AC modulation of the strength of these ionospheric currents is relatively new. The amplitude-modulated electrojet current radiates AC signals both down to the Earth's surface and out into the magnetosphere. Ionospheric conductivity may be modulated naturally, for example by

modulating the flux or energy of the auroral electrons during a pulsating auroral event or by pulsations of the geomagnetic field. It may also be artificially modulated by man-made radio signals such as the VLF Omega signals between 10 and 14 kHz used for long range navigation, or signals from low or medium frequency radio transmitters (see Cannon, P.S., Turunen, T. & Rycroft, M.J. *J. atmos. terr. Phys.* **44**, 831; 1982). It can also be achieved by using an HF transmitter, purpose-built at high latitudes, that acts as an ionospheric heater. All these examples are believed to be explained by a non-linear plasma physics mechanism: the energy of the incident radio wave heats the ionosphere and increases the electron temperature, thereby causing a modulated variation of the collision frequency between ionospheric electrons and neutral atoms. This, in turn, causes an AC variation of the ionospheric conductivity. In the presence of the large ambient DC current — measuring millions of amps — this current is also modulated. As a result, it acts as an enormous transmitting antenna placed in the lower ionosphere.

The new results reported by Barr *et al.* derive from operating an ionospheric heater, a transmitter radiating at 2.7 MHz, which is amplitude-modulated at ELF, near 1 kHz. Situated at Tromsø, Northern Norway, it lies below the auroral electrojet. Measurements of the signal at the frequency of the amplitude modulation were made nearby, and also at distances of some 200 and 550 km. This ELF signal is produced in the ionosphere, by modulating the conductivity of the auroral electrojet, which acts as a large transmitting antenna effectively exciting the Earth-ionosphere waveguide. The temporal variability of these signals, observed at the two remote sites, is somewhat similar, so that variations in the strength of the source, rather than variations in the propagation conditions, are needed to explain the variability. The variations are found to be partly associated with variations in the geomagnetic activity with which the electrojet itself is associated. The efficiency of this antenna, in terms of the radiated ELF power in relation to the power of the high frequency (HF) heater, lies between 10^{-12} and 2×10^{-9} . The derived attenuation rate for ELF propagation in the Earth-ionosphere waveguide is 5–30 dB/Mm.

Another recent paper (Papadopoulos, K. & Chang, C.L. *Geophys. Res. Lett.* **12**, 279; 1985) shows that amplitude-modulated HF heating of the ionospheric E region can produce ULF/ELF signals. These signals are considered to be associated with the spontaneous generation of alternating magnetic fields caused by the coupling of the 'hot-spot' temperature gradient to the ambient ionospheric plasma-density gradient. This mechanism, which will be effective in the absence of large ionospheric currents, is considered theoretically to have an efficiency of 10^{-5} .

For the future, there is much to be done in improving our understanding of both the plasma physics and the conditions under which such amplitude-modulated signals may be broadcast by the ionosphere. A further type of experiment in a related field involves the space shuttle. An instrumented package of about 1.5 m diameter can be tethered to the shuttle by a conducting wire some 20 km long, which is deployed from a Spacelab pallet mounted on the shuttle. The long wire antenna could act as a transmitter of waves with

frequencies ranging from ULF to VLF when an electron gun operating on the space shuttle itself is switched on and off at such frequencies. NASA is planning to test this concept in a few years. This exciting experiment, which should produce observable effects both on the ground and in the magnetosphere, is just one of several possible uses of a sub-satellite tethered to the space shuttle. □

M.J. Rycroft is Head of the Atmospheric Sciences Division, British Antarctic Survey, NERC, Madingley Road, Cambridge CB3 0ET, UK.

Ocean Drilling Program

Evolution of a passive margin

from the Leg 103 shipboard scientific party*

LEG 103 of the Ocean Drilling Program was devoted to unravelling the tectonic and sedimentary evolution of the Atlantic margin of the Iberian Peninsula. The transect of five sites, with a total of 14 drill holes, at the seaward edge of the Galician margin revealed a complex history of subsidence and rifting preceding the initiation of seafloor spreading between Newfoundland and Iberia. The drilling encountered a great variety of sedimentary rock types, in addition to a suite of volcanic and metamorphic rocks. Several unexpected discoveries illustrate the importance of obtaining data by deep drilling to provide constraints for geophysical models of the rifting process.

The main findings include the following: (1) a ridge of material of mantle composition is located near the boundary between oceanic and continental crust, (2) the 'basement seismic reflector' beneath several tilted continental blocks is formed by a Mesozoic shallow-water carbonate platform, (3) platform drowning, tilting of fault blocks and rapid subsidence preceded the first formation of oceanic crust by as much as 25 million years, (4) there was abundant influx of terrigenous clastic sediments to deep-water submarine fans during the early stages of rifting, and (5) the seismic reflector does not, as widely believed, represent a ductile-brittle boundary within the continental crust but is instead a reflector at the base of the sediments deposited during the rifting stage.

It is debatable whether genuine continental basement was drilled during the leg. The oldest rocks recovered at the base of the main site are several types of slightly metamorphosed sedimentary and volcanic rocks, but these are interpreted either as a basal conglomerate or scree on a basement subcrop beneath Cenozoic and Mesozoic sediments, or perhaps as *in situ* volcanic rock, forming part of the basement. It was not possible to drill deeper to explore the geology of the basement. The apparent absence of sedimentary rocks older than late Jurassic (about 150 million

years before present) contrasts with the thick marine and clastic deposits in basins in Portugal and off Newfoundland, which record pulses of subsidence and rifting beginning in the Triassic (230 Myr).

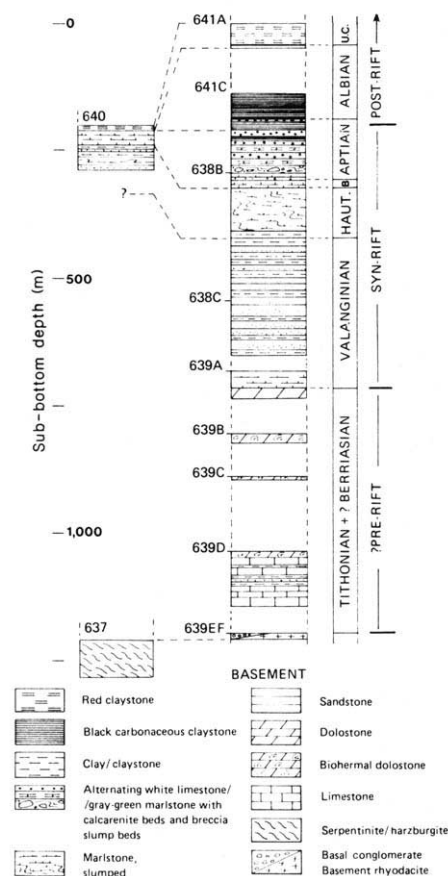
A shelf and carbonate platform environment prevailed at the site of the future margin during the late Jurassic and earliest Cretaceous (150–140 Myr). Approximately 400 m of limestone were deposited, of which the upper part is extensively altered to dolomite. The top of the well-cemented carbonates forms a prominent seismic reflector which effectively conceals the deeper strata and basement on seismic reflection profiles. This seismic reflector had earlier been interpreted as 'basement'. Near the foot of the margin, a strong, smooth reflector occupies a similar position beneath the overlying thick sequence of Lower Cretaceous turbidites. This reflector has been postulated to be a brittle-ductile transition level or a horizontal shear plane within the continental crust, into which the deep bounding faults of half-grabens merge. The new drillings therefore cast doubt on current geophysical models of passive rifted margins based on seismic profiles in this region and adjacent margins.

The carbonate platform drowned near the beginning of the Valanginian stage (135 Myr) is the first episode of rapid subsidence recorded at the drill sites. Sediments deposited since this drowning consist of as much as 3 km of pelagic carbonate and clay, and of several intervals enriched in clastic detritus. In general, subsidence was more rapid than sedimentation; the average depth of the ocean floor at the drill sites is now about 4,500 m.

Above a thin interval of Lower Valanginian marl, thick wedges of deep-water terrigenous clastics partly filled active fault basins on the margin. The drill ship explored only a thin edge of one of these wedges, but the interpretation of seismic profiles shows that in less than 15 million years as much as 1,500 m of sand and silt were deposited in the nearby grabens by turbidity currents and debris flows,

perhaps associated with submarine fans. The source of this clastic detritus may have been Newfoundland, Iberia or an unidentified land mass to the north, or a combination of these. Many beds are rich in wood and other terrestrial plant fragments. Similar Early Cretaceous submarine fan systems have been discovered by drilling off other North Atlantic margins, although their exact timing and duration differ from place to place. The main episode of terrigenous sand deposition at the sites drilled on Leg 103 terminated in the early Hauterivian (130 Myr). Sediments of late Hauterivian to early Aptian age (130–115 Myr) consist of alternating bioturbated limestone and laminated marl—a pelagic facies typical of oceanic sediments of this age throughout the North Atlantic. These sediments contain abundant slumps, debris flows and thin clay-rich turbidites, indicative of slope deposition and synsedimentary faulting. The seismic profiles indicate that nearby half-grabens underwent deepening or sediment filling, or both.

Seismic and drilling data suggest that block faulting began at least as early as the Valanginian and continued intermittently until the end of the Aptian. The last intense faulting associated with rifting is tentatively assigned to the latest Aptian (110 Myr). This episode is marked at the Leg 103 drill sites by the termination of



Composite stratigraphic columns of sites drilled on ODP Leg 103. B, Barremian; U.C., Upper Cretaceous.

influx of shallow-water carbonate debris, a rapid reduction in the amount of pelagic carbonate and a period of slow or condensed sedimentation; it is followed by the onset of accumulation of carbonate-poor bioturbated clay containing organic-rich layers (black shales). The event is marked regionally by a prominent seismic reflector and local unconformity, and seems to coincide with the initiation of seafloor spreading between Iberia and Newfoundland as determined from magnetic anomalies.

Drilling also confirms the presence of serpentinized peridotite, which is probably altered rock from the upper mantle, on a long ridge close to the boundary between oceanic and continental crust. This ridge might be a fragment of Palaeozoic (300–500 Myr) basement, or it may have

been emplaced during the last stages of the rifting process as a result of thinning of the crust. The changes in sedimentary regime accompanying this rifting are very clearly documented by the drilling results from Leg 103, which have also revealed the hazards of geologic interpretation based solely on geophysics. □

*Co-chiefs scientists: G. Boillot (Univ. Pierre et Marie Curie, France) and E. L. Winterer (Scripps Institution of Oceanography). Also J. Applegate (Florida State Univ.), M. Baltuck (Tulane Univ.), J. A. Bergen (Florida State Univ.), M. Comas (Univ. de Granada, Spain), T. A. Davies (Univ. of Texas, Austin), K. Dunham (Univ. of Michigan), C. A. Evans (Colgate Univ.), J. Girardeau (Institut de Physique du Globe, France), D. Goldberg (Lamont-Doherty Geological Observatory), J. Haggerty (Univ. of Tulsa), L. F. Jansa (Bedford Institute of Oceanography, Canada), J. A. Johnson (Univ. of California, Los Angeles), J. Kasahara (Earthquake Research Institute, Univ. of Tokyo), J.-P. Loreau (Museum National D'Histoire Naturelle, Paris, France), E. Luna (Hispanoil-Enipsa, Madrid, Spain), A. W. Meyer (ODP, Texas A&M Univ.), M. Moullade (Univ. de Nice, France), J. Ogg (Scripps), M. Sarti (Univ. di Ferrara, Italy), J. Thurov (Univ. Tübingen, F.R.G.), M. W. Williamson (Dalhousie Univ.).

Supernovae

Little Bear's mass loss rate

from Virginia Trimble

WHICH STARS explode as supernova, and how do they do it? Astronomers have debated these questions since Walter Baade and Fritz Zwicky identified and defined supernovae as a distinct astronomical phenomenon in 1934¹. Three recent contributions^{2–4} seem to have resolved at least one facet of this problem. The current consensus is that Type I (hydrogen free) events arise in relatively old, low-mass stars and derive energy from explosive burning of carbon and oxygen. Type II (hydrogen-rich) supernovae, on the other hand, end the lives of young, massive stars, whose cores collapse to neutron star densities or beyond. The light we see comes from a shock wave propagating through many solar masses of hydrogen in an envelope 10^{14} – 10^{15} cm across. Massive stars develop such envelopes during their late, red supergiant evolutionary phases.

Without these envelopes, broad-peaked light curves cannot be reproduced by the models^{5,6}. And herein lies the main discrepancy with the present majority view that all or most single and wide binary OB stars shed their extended, hydrogen-rich envelopes, becoming, first, Of stars, and then Wolf-Rayet (WR) stars before exploding^{7–9}. There could still be enough hydrogen (0.1 – $0.3 M_{\odot}$) to produce the spectra's lines we see, but the exploding stars will be compact. As a result, most of the shock energy goes into expanding the star rather than into radiation, yielding a very dim supernova. And if there is too little hydrogen in the envelope, its recombination cannot produce the prolonged plateau phase seen in most Type III light curves.

The first of the new contributions comes from Schild and Maeder², who have looked again at the numbers of OB stars

versus numbers of WR stars. They conclude that only stars in excess of $18 M_{\odot}$ give rise to nitrogen-rich WRs and only those above $35 M_{\odot}$ make the carbon-rich ones. In addition, they find that most 18 – $40 M_{\odot}$ stars need never go through a WR phase at all. This permits a reasonable range of intermediate mass stars to retain their extended, hydrogen-rich supergiant envelopes and make Type II light curves and spectra. One implication is that if the progenitor of the Crab Nebula, at $9 \pm 1 M_{\odot}$, stripped most of its envelope before exploding, it was unusual in doing so. This is good; it helps explain the rarity of pure, filled-centre remnants.

Second, Niemala, Ruiz and Phillips³ have, through a combination of luck and preparedness, managed to catch the Type II event 1983k in NGC4699 nearly 10 days before maximum light. The pre-maximum spectrum shows N III and He II emission lines superimposed on a strong blue continuum, suggesting a progenitor whose surface composition resembled that of a WR star. Near maximum light the emission lines disappear, leaving weak H I, He I and Ca II absorption lines. A month after maximum, the spectra are dominated by the P Cygni emission lines of H I that characterize normal Type II supernovae. The light curve is also quite normal. The implication is that we were seeing, first, emission from photospheric material as the shock wave emerges from the stellar core, then absorption in a normal-composition circumstellar envelope (shed by the star prior to core collapse), and finally, emission lines from the hydrogen-rich zone when the shock reaches and heats it. The light curve has a very broad peak, implying a very extended, pre-existing circumstellar shell.

Niemala *et al.* conclude that 1983k most likely resulted from a massive star that retained five or more solar masses of hydrogen-rich envelope but had also lost 1 – $2 M_{\odot}$ into a circumstellar shell before exploding. The surface layers contained nitrogen-rich material, but the authors do not insist that the star would have looked like a WR if observed pre-explosion. They consider it equally possible that the WR-like lines they see are a result of the physical conditions produced by the supernova event itself. One last surprise: supernova 1983k occurred 17 ($75/H$) kpc from the centre of its parent galaxy, in projection, and well away from any H II regions or spiral arms that could be seen in a deep charge coupled device plate of the galaxy. The progenitor must, therefore, have been a runaway star or have been formed well away from the usual spiral arm sites.

The third contribution is from Josef Shklovskii, who tied several ideas together shortly before his death to provide an explanation of the absence of Type II supernovae in Magellanic irregular galaxies⁴. This absence has long been a puzzle, because these galaxies are the richest of all in gas and young stars. Yet only Type I supernovae have been detected in them. Shklovskii's explanation is that massive stars in these galaxies never develop the extended stellar or circumstellar envelopes needed to make Type II light curves as a direct result of their low metal abundances. Radiation passing through a star needs the absorption lines of heavy elements in order to transfer momentum to the outer layers and drive them off in a wind. Thus, massive stars in irregular galaxies never expand properly and, like stars that have been completely stripped, produce very dim supernovae. Shklovskii noted the progenitor of Cas A as a possible example of nonexpansion in our own Galaxy. I suspect complete stripping is also a possible interpretation for the faintness of that explosion.

The net conclusion we can draw from these studies^{2–4} is that the ideal Type II progenitor must have just the right amount of mass loss. Such stars occur in the Milky Way, NGC4699, and other spiral galaxies, but do not occur in elliptical or irregular galaxies. □

1. Baade, W. & Zwicky, F. *Proc. natn. Acad. Sci.* **20**, 259 (1934); *Phys. Rev.* **45**, 38, 46, 76 (1934).
2. Schild, H. & Maeder, A. *Astron. Astrophys.* **136**, 237 (1984).
3. Niemala, V.S., Ruiz, M.T. & Phillips, M.M. *Astrophys. J.* **289**, 52 (1985).
4. Shklovskii, J.S. *Sov. Astron. — AJ Lett.* **10**, 302 (1984).
5. Falk, S.W. & Arnett, W.D. *Astrophys. J. Suppl.* **33**, 515 (1977).
6. Woosley, S.E. & Weaver, T.A. *Preprint* (proceedings of 1983 Santa Cruz workshop) (1985).
7. Conti, P., Garmany, C., de Loore, C. & Vanbeveren, D. *Astrophys. J.* **274**, 302 (1983).
8. Falk, S.W. & Mitalas, R. *Mon. Not. R. astr. Soc.* **202**, 19 (1983).
9. Bertelli, G., Bressan, A.G. & Chiosi, C.S. *Astron. Astrophys.* **130**, 278 (1984).

Virginia Trimble is Professor of Physics at the University of California, Irvine, California 92717, USA; and Visiting Professor of Astronomy at the University of Maryland, College Park, Maryland 20742, USA.

Hazards beyond number

David L. Sills

Risk Watch: The Odds of Life. By John Urquhart and Klaus Heilmann.

Facts on File: 1985. Pp.214. \$16.95, £11.95.

Risk: Man-made Hazards to Man. Edited by M. G. Cooper.

Clarendon: 1985. Pp.141. Hbk £15, \$23.95; pbk £9.95, \$12.95.

HERE are two readable, informative and in the main "right-minded" books on the subject of risk in a technological society. By "right-minded", no more is meant than that I share their basic attitude towards this highly-controversial topic. Both books assert that on balance life is safer today than it has ever been in the past; that is, neither of them is a "doomsday" tract. Both emphasize such determinants of risk as technological complexity, human error and statistical probability, and avoid singling out evil industrialists or incompetent bureaucrats as the main creators of risk. Both assert that there is no turning back to a romanticized pastoral, craft-dominated, risk-free society. But both present a somewhat oversimplified version of how things can go wrong, of how we are exposed to risks. I will turn to the question of what they leave out in a moment.

Risk Watch has an interesting publishing history. The authors are physicians and by nationality American (Urquhart) and German (Heilmann). Urquhart works for a Californian high-technology firm specializing in drug-delivery systems; Heilmann is a professor of ophthalmology at the Technical University of Munich and a consultant in medical engineering at the University of California, Los Angeles. The book was first written in English and German, and published in German in 1983 as *Keine Angst vor der Angst* (*No Fear in the Face of Fear*), with Heilmann as the senior author. The present English-language edition is a translation and revision of the German edition.

The polemical aim of *Risk Watch* is anti-alarmist. "Risk-taking and risk-avoiding

are intrinsic to life itself", runs the first sentence of the introduction. The authors believe that the media, relying on "victim-oriented" and other forms of dramatic reporting, overemphasize some risks (nuclear energy, air travel) and under-emphasize others (cigarette smoking, automobile accidents). Accordingly, they also have a didactic aim: to inform the general public, towards whom the book is addressed. Like others before them, the authors affirm their faith in public enlightenment by quoting Thomas Jefferson on the "ultimate powers" of society residing in the people, and the consequent necessity "to inform their discretion".

Because Urquhart and Heilmann are physicians, *Risk Watch*, appropriately enough, has chapters devoted to medical/surgical risks and risks from food consumption. The emphasis throughout is upon individual risk-taking behaviour and aggregate statistics. A "Safety-degree Scale", developed by the authors and based on the frequency of events and the number of people exposed to these events, rates cigarette smoking and motorcycling as high risks; oral contraceptive pills and bicycling as medium risks; and bee stings and being hit by falling aircraft as low risks. The authors hope that their scale will do for risk reporting what the Richter Scale has done for earthquakes: enable the media to use a scientifically based scale in their news stories on accidents, calamities and other events in a risky world.

Risk: Man-made Hazards to Man is based on eight lectures given at Wolfson College, Oxford, in the spring of 1984.

The lecturers are scientists and public officials: six of the eight are professors at Oxford or other British universities, as is M. G. Cooper, the editor, who teaches engineering at Oxford. The style of presentation is informal but informed. A chapter each is devoted to medical intervention (W. H. W. Inman), biological research (G. E. Gordon-Smith), industrial risks (H. J. Dunster), industrial chemicals (A. E. M. McLean), the energy industry (Frederick Warner) and environmental change (Richard Southwood).

Each of the chapters reviews what is known about the risks of one sector of modern life, and the authors' conclusions are generally sanguine. We are told by Inman that "medical intervention with modern drugs is remarkably safe" (p.52); by Gordon-Smith that risks both to workers and the public from biological research are "small" (p.73); by Dunster that "direct industrial risks to life and health do not threaten our society" (p.93); by McLean that "our use of industrial chemicals has not had the disastrous effects that people once feared" (p.108); and by Warner that "the risks involved in energy production . . . have been shown to be extremely low, and not to require more provision for their reduction" (p.124). (Warner uses only British data on coal production, where there indeed has been a marked decrease in deaths and injuries in the past 30 years, but there are thousands of newly bereaved coal-mining families throughout the world who could not accept his Pollyanna-ish conclusion. He is not alone in this seeming blindness: in polite society, coal-mining and coal-burning accidents happen to somebody else.) Only Southwood sounds a quiet alarm, concluding that environmental changes, "in their universality and in the extent to which there is no recent precedent, dwarf all other man-made effects" (p.132).

D. E. Broadbent's introductory essay, on the psychology of risk, is the only contribution in the book by a social scientist; the editor weakly apologizes by stating that the social sciences "could not be given

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

adequate representation within the overall limit of eight lectures" (p.1). Similarly, the authors of *Risk Watch* fail to recognize that risk is only in a very limited sense a statistical concept. At its heart, risk is essentially a culturally learned response to uncertainty, a point made repeatedly by Mary Douglas and Aaron Wildavsky in *Risk and Culture* (University of California Press, 1982), a book that should be read in conjunction with the two under review. Hermann Bondi, then of the Natural Environment Research Council, in a perceptive introductory essay in *Risk*, captures the contradictions inherent in risk assessment by noting that people today want to be sure that "at the end of their day's work they can get away safely to go hang-gliding" (p.12).

Neither Urquhart and Heilmann nor the contributors to *Risk* seem to recognize that the identification and management of risks are political problems (the "politicizing of nature" is the term used by Douglas and Wildavsky). For example, although *Risk Watch* was first published in Munich, it makes only passing reference to the Greens, the West German political party that is the only successful environmental party in a major industrial nation. The programme of the Greens is based on the premise that large-scale technology and competitive capitalism are destroying the environmental foundations of human life; its assessment of the nature of risk is in sharp contrast to that of these authors.

Risk is a concept that varies in meaning from culture to culture. The authors of these two books largely ignore this. They also ignore the point that the importance placed upon risk and its determinants is a political fact of our time: witness, for example, the various interpretations of the recent tragedy at the Union Carbide plant in Bhopal. And they do not recognize that accidents cannot be adequately explained by the interaction between a flawed technology and statistical probability; on the contrary, accidents happen in large part because of flaws in the social system that controls the technology. This was the main conclusion of all the inquiries into the 1979 accident at the Three Mile Island nuclear power plant and was the theme of one of the scholarly by-products of that accident, Charles Perrow's *Normal Accidents* (Basic Books, 1984), which I reviewed for *Nature* last year (309, 185; 1984).

Both books, then, give short shrift to the influence of culture, politics and social systems on risk and its perception. They are nonetheless important counterweights to journalistic sensationalism and useful introductions to the burgeoning field of risk-benefit analysis. □

David L. Sills is Executive Associate at the Social Science Research Council, 605 Third Avenue, New York, New York 10158, USA. He is a co-editor of Accident at Three Mile Island (Westview, 1982) and The Three Mile Island Nuclear Accident (New York Academy of Sciences, 1981).

Some information on recombination

Harold L.K. Whitehouse

Genetic Recombination. Edited by John H. Wilson. *Benjamin/Cummings*: 1985. Pp.517. \$29.95, £29.95.

TEN articles from *Annual Reviews of Genetics*, four from *Annual Reviews of Biochemistry* and one from *Annual Reviews of Plant Physiology* make up this volume. The articles were originally published at various dates from 1975 to 1983 and are reprinted unchanged, but with new page numbers. All are concerned either directly or marginally with aspects of genetic recombination, and all of the authors have addresses in the United States or Canada.

The reviews are grouped under three headings: meiotic, mitotic and bacterial recombination. This division is somewhat arbitrary, as J.H. Wilson states in his preface to the book. Thus, the meiotic section includes F.W. Stahl's review of 1979, which is based primarily on the chi sites in bacteriophage λ of *Escherichia coli*, and C.M. Radding's review of 1978, which is concerned with molecular studies and so necessarily centres on work with prokaryotes. In the mitotic section there is an article by R.A. Weinberg on animal virus genome integration, and one by H.O. Smith and D.B. Danner on genetic transformation which, not surprisingly, is largely concerned with bacteria. The bacterial section is also misnamed as two of the reviews deal with bacteriophage recombination. A more serious criticism of this grouping is that it tends to obscure a central issue, namely, the need to see recombination phenomena as a whole. Many of the reviews are relevant to both prokaryotes and eukaryotes and, within the latter, to both meiotic and mitotic events. Putting the reviews into categories serves little purpose, though the helpful index does mitigate the problem to some extent. In a rapidly advancing field a chronological sequence might have been best.

One of the most useful papers in the book is a detailed survey by B.S. Baker *et al.* of what is known about the genetic control of meiosis in fungi, *Drosophila* and flowering plants, supported by nearly 700 references. There are also two valuable accounts of molecular studies on recombination, one by C.M. Radding and the other by D. Dressler and H. Potter. On the other hand, several of the reviews are concerned with topics that have contributed little to understanding the mechanisms of recombination, for example, sister chromatid exchange (reviewed by S.A. Latt); animal virus genome integration; and the molecular genetics of bacteriophage P1 (N. Sternberg and R. Hoess).

Although the publisher states on the cover that the volume represents the best overview available today, this is not to say that the reader will obtain a current picture of the subject. There is no review of transposable DNA and no discussion of the double-strand-break repair model of recombination proposed in 1983 by J.W. Szostak and associates. A postscript to bring the topic up to date would have improved some of the older reviews, for example Stahl's on specialized sites in recombination, that by P.J. Hastings on fungal recombination and that by T.R. Broker and A.H. Doermann on recombination in bacteriophage T4. Better than nothing would have been a list of references to important new work. □

Harold L.K. Whitehouse is Emeritus Reader in Genetic Recombination in the Department of Botany, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

Talk of the town

Margaret C. Jacob

William Whiston: Honest Newtonian. By James E. Force. *Cambridge University Press*: 1985. Pp.208. £25, \$37.50.

TRY TO conjure up a London coffee house of the early eighteenth century, where the genteel and the mercantile jostle with clergymen as well as fellows of the Royal Society; and then ask to listen in on their discourse. That, in effect, is what cultural historians with an interest in the reception of science in the time of Newton have been doing for the past 15 years or so. We have now succeeded, through a reading of the printed and manuscript sources of this period — when Newton's science and the recently published *Principia* (1687) were first discussed — to reconstruct in a plausible way what those gentlemen were saying (women did not frequently attend coffee houses but they did go to hear sermons).

James E. Force's book on William Whiston, Anglican clergyman, pulpit orator, coffee-house lecturer, is a fine contribution to that reconstruction and it comes with the added claim that Whiston, one of Newton's earliest and closest followers, more accurately and honestly (than other Newtonians) said what the master truly believed. For example, Newton and Whiston were anti-Trinitarians; only Whiston went out on the limb and much to the horror of his mentor said it publicly. As a result, he lost his professorship in Cambridge. In addition both were avid millennarians, that is, they believed in the end of historical time as foretold in the Bible, a destruction and consequent paradisiacal reconstruction to be accomplished by direct divine intervention. To put it simply, a modern thinker could hear totally unexpected discussions in those coffee houses and pulpits, indeed

be shocked by the total absence of a distinction between science and religious belief, between devotion to mathematics and a literal reading of the Bible.

As Force amply documents, Whiston (as did Newton, incidentally) belonged totally to that world; yet 15 or 20 years ago scholars barely understood it. In possession at that time of well-formulated modern distinctions between science and religion, they simply ignored Newton's manuscripts on the apocalypse or dismissed Whiston as a crank. It is not immodest to say that in *The Newtonians and the English Revolution* (Cornell University Press, 1976) I was one of the first to argue for a literal reconstruction of early-eighteenth-century discourse, and in many places Force is in dialogue, generally a sympathetic one, with that book. In general I am delighted by the astuteness of his observations, although one question must be addressed.

What is the general reader to make of a book such as this? It is beautifully researched and nuanced, scrupulous in its refusal to apply modern categories to the early modern mind of either its subject or Newton himself — in short, a model of the serious new history of culture that can now be written. Yet it remains a monograph, narrowly focused, intent upon reconstructing the religious thought of Whiston, and indeed the millenarianism of Newton, resolutely uninterested in the larger world of church preference or the social uses to which Newton's ideas were put by the early-eighteenth-century church. Indeed, at moments scholars such as myself are taken to task for painting on too large a canvas, for being too general.

Yet such books must be written and read. They advance scholarship through the sheer depth of their research; they make us feel as if we are eavesdropping on one of those early-eighteenth-century conversations. For that reason the general reader should peruse Force's book and similar works. How else can we recapture a world where the founder of modern science believed that God would destroy that very natural world explicated in the *Principia*, who also urged his followers, as Force shows, to preach both his science and his religion from the pulpit and in the end, given the reality of his desire for power, repudiated Whiston because he said too much in print as well as in those coffee houses? Books like Force's can humble us before the extraordinary capacity of the past to refuse to be like the present, before its ability to violate our most cherished assumptions about the nature of modern science, not least its very origins. □

Margaret C. Jacob is University Professor of History on the Graduate Faculty of the New School for Social Research and Dean of Lang College, NSSR, 66 West 12th Street, New York, New York 10011, USA. Her most recent book, *Science and Elites: The Integration of Science in Western Culture*, will be published next year by Knopf.

The dangers of being discovered

Tim Halliday

Immigrant Killers: Introduced Predators and the Conservation of Birds in New Zealand. By Carolyn King. Oxford University Press:1985. Pp.223. £29.50, \$34.95.

OVER the past few hundred years, the pattern of largely isolated communities of endemic plants and animals that was the legacy of continental drift and island formation has been radically altered, in some places destroyed, by human colonization, which invariably involves the introduction of a variety of animals to areas where they previously did not occur. Introduced animals frequently are serious predators, parasites or competitors in an indigenous fauna that has had no opportunity to evolve effective defences against them. While European starlings, sparrows and rats now flourish in many parts of the world, many birds that previously lived on isolated islands have become extinct.

Nowhere is this process more apparent than in New Zealand, where once vast forests have been reduced to relic patches of woodland in an agricultural landscape deliberately intended to resemble that in Britain, complete with European deer, rabbits, weasels and rats. Carolyn King focuses on predatory species that have been introduced to New Zealand, since it is often they that have been blamed for the demise of many of New Zealand's endemic birds. In a thorough and well-researched account of the history of its fauna and flora, she evaluates all the factors that have contributed to the upheaval in the ecology of New Zealand that began with colonization by Polynesians about 1,200 years ago, so that the effect of introduced predators on contemporary populations of endemic birds can be accurately assessed.

What emerges is that the impact of

weasels, stoats, cats and other predators is appreciable, but relatively minor compared to direct human activities such as deforestation and hunting. The author also emphasizes the intrinsic vulnerability of New Zealand's endemic fauna, a result of its long history of geographical isolation. Until human beings introduced them, there were no indigenous mammals, and animals of different kinds evolved to fill the niches occupied by mammals elsewhere; the weta, a giant cockroach, is the equivalent of a mouse and the kiwi is like a rat in many of its habits. A general feature of New Zealand's animals, like those on islands throughout the world, is their lack of defences against mammalian predators. To varying degrees, many birds lost their powers of flight and became pathetically unafraid of human beings and their fellow immigrants; both European settlers and the Polynesians that preceded them were able to slaughter them with ease. Of more significance, however, was the reckless and relentless destruction of the native forests to make way for agricultural land. One would like to think that such catastrophic practices would not be countenanced in today's more ecologically conscious world but, as Carolyn King describes, where there are powerful economic forces at work, natural habitats are probably never secure.

This book is well-written and beautifully illustrated; some of the photographs of the destruction wrought by early European settlers are particularly striking. The tone of the book is inevitably somewhat gloomy but the author avoids being self-indulgent in her sense of loss for the wildlife of her adopted country. If the human species is to avoid extinction at its own hands, it needs to be constantly reminded of its past follies and excesses. The work and writings of biologists such as Carolyn King provide such a reminder and should be heeded by those who have any concern at all for the future of life on Earth. □

Tim Halliday is Reader in Biology at the Open University, Walton Hall, Milton Keynes MK7 6AA, UK.



On the rampage in New Zealand — hunters use one introduced species, the dog, to chase another, the pig. (An engraving from *The Illustrated New Zealand Herald*, 1868.)

Devil of a business

Anthony Sampson

Fundamentals of the Petroleum Industry.

By Robert O. Anderson. Weidenfeld & Nicolson/University of Oklahoma Press: 1985. Pp. 390. £18.95, \$29.95.

ROBERT ANDERSON'S *Fundamentals of the Petroleum Industry*, lavishly printed and illustrated, is clearly designed to be a standard guide to the oil business — a successor to an earlier book of the same title by Dorsey Hager. Much of it is a description, in clear and upretentious language, of the basic engineering and organization of the oil business, including the geology, drilling, transport and refining. But it has the added interest of being written by one of the most successful, and wealthiest, practitioners in the industry. Anderson made a fortune as an oil entrepreneur in New Mexico, which he then extended as chairman of Atlantic Richfield based in Los Angeles.

Atlantic Richfield (Arco) has been one of the most aggressive and well-managed rivals to the traditional multinational oil companies, the "seven sisters"; and Anderson himself claims (with some exaggeration) that Arco is "often regarded as the Eighth Sister". But certainly the company has had a remarkable record in exploration and development — above all in

finding oil in Alaska — which adds authority to the book's technical chapters.

Anderson is financier as much as an oilman, and knows full well that oil is increasingly a business "of partnerships and shareholders, bankers and backers". So while the book provides a readable beginner's guide to the industry's workings, complete with diagrams, graphs and a glossary, it also includes financial and political judgements which have a wider interest.

Predictably Anderson is a champion of free enterprise in oil: governments, he says, "do not have the pressures, the energies, the tempo, the dynamics of the private sector"; the speed of drilling in Russia, for instance, is five to ten times slower than in North America. He explains with natural certainty that in the oil business "there was, and is, still room for individuals with friends, faith, perseverance, some money, and a lot of luck to become independently wealthy", but is less convincing when he asks: "who then owns the petroleum industry? Its customers and employers do, in a very real sense. And who pays for it? The very same". In similar vein he is critical of many government attempts to regulate and control his industry, which he sees as the only one with the resources to develop alternative energy sources.

Anderson, however, is also a realist: he recognizes that outside the United States governments will insist on having their own oil corporations and controls, and concedes that even in America the "invisible hand" of free enterprise may not be enough: "if even the gentlemanly game of tennis must have umpires, referees and linesmen, then how much more so must the game of capitalism have them, where the stakes are so high?".

The early chapters provide a rapid history of oil since its discovery in Pennsylvania, re-telling familiar stories about wildcatters, gamblers and monopolists with some marvellous old photographs but no obvious new research. Rather, it is the assessment of the recent past and the future which is most useful.

Anderson points out that "learned people have been predicting the exhaustion of the world's oil since 1863" and how "the day of reckoning has always seemed to be about twenty years in the future. Somehow, a new series of major discoveries has always put off that day". Will that postponement continue? The forecasts remain very divergent, both about future production and consumption. For the time being, as Anderson describes it, the Organization of Petroleum Exporting Countries (OPEC) can fix a price which is low enough to keep most new alternative energy — whether tar-sands, solar energy or windmills — uncompetitive with oil. But most experts, including Anderson, believe that before long oil shortage will push up the oil price above the equivalent for new forms of energy:

At the crossover time, probably in the 1990s [says Anderson], the combined productive capacity of all conventional sources, including OPEC, will no longer equal demand at a price OPEC can influence. Demand will pull prices up, and no producer will have the power to depress them.

Forecasts of the price of oil have been notoriously unreliable: even since Anderson finished this book — two years ago; it was published in the United States in 1984 — the Western consumption of oil has diminished and new oilfields have been found, pushing back the likely date of that crossover. The day of reckoning may still be about 20 years ahead, but it cannot be postponed much further.

The future control of oil is very uncertain. Mr Anderson's historical chapters



1973 — time of another oil crisis.

vividly describe the extraordinary alternations of glut and shortage in the past, which caused crazy oscillations in the oil-price until production was effectively controlled — whether by Rockefeller, or by the Texas government, or by the Seven Sisters, or by OPEC. For the foreseeable future he believes (or believed in 1983) that OPEC will be "a major influence on world supplies and world market prices". Other oil-producers outside OPEC will consume all they produce. But the new era of alternative energy should produce a world market which will be less dependent on the notoriously slippery behaviour of the oil-price.

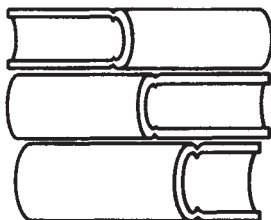
Or will it? Anderson's clear exposition of the remorseless logic of world energy can be misleading. For behind all that intricate engineering and global organization oil has kept on playing tricks which have often made fools of experts and bankrupted rational corporations. As the Mexicans say: "God gave us the land, the devil gave us the oil". □

Anthony Sampson, 27 Ladbrooke Grove, London W11 2AY, UK, is author of *The Seven Sisters: The Great Oil Companies and the World they Shaped* (Hodder & Stoughton, 1975).

MONOGRAPHS TEXTBOOKS...

and other technical
publications
on Nuclear Energy

write for free
publications and
services catalog



American Nuclear Society
555 N. Kensington Ave.
La Grange Park, IL 60525
(312) 352-6611



Religion and biology beyond conflict

George M. Marsden

Darwinism and Divinity: Essays on Evolution and Religious Belief. Edited by John Durant. Basil Blackwell: 1985. Pp.210. £15, \$24.95.

ONE of the main intellectual trends of the past 25 years has been the growing sensitivity to the sociology of knowledge. Our most carefully reasoned views, we are constantly being reminded, are shaped by complex affiliations and considerations that seem to have little to do with our reasoning itself. The impact of these extra-rational factors has long been recognized in areas involving soft reasoning, such as political thought. The revolution of the past 25 years is that such factors have come to be widely regarded as playing a large part in even our hard empirical reasoning. A host of post-Kuhnian historians of science have reinforced this point. Natural science, they argue, takes place not in the stable environments that we might imagine, but (to borrow an image from Richard Lewontin) as though bouncing back and forth on multiple trampolines of cultural forces that it both reshapes and is redirected by.

Nowhere are the reciprocal interactions of scientific and non-scientific cultural forces more evident than in the history of Darwinism. Most centrally, Darwinism has always aroused passions because it has been so mixed up with religion or the lack thereof. In addition, Darwinism and divinity do not even exhaust the principal forces in the interplay. As the present volume shows, both of them constantly have been used to promote or attack various social-political programmes. So while modern biology is shaped largely by a valid internal logic, its repeated interactions with these two most emotionally laden cultural activities, religion and politics, have long made it difficult to keep a clear focus on simply the biological issues.

Darwinism and Divinity is a series of essays which has grown out of a conference organized by the British Society for the History of Science in 1982. The works here selected reflect the rapidly growing consensus of the past decade that the old conflict paradigm for understanding relations between religion and science, even biological science, simply fails. John Durant, the editor, describes the volume as a kaleidoscope of knowledge in which each chapter reshapes the subject of religion and biological theory and displays new sets of relationships. If there is a coherence to such a design, it is in the underlying principle that the subject of biological thought and divinity is more complex than most people realize.

Whether by accident or design, the

essays display a number of other themes. Most central is that biological evolution constantly has been used illicitly to support differing views of the world. In his survey of a century of debate over Darwinism and religion, Durant speaks of the "idolatry" of the "deification of evolutionary process". Mary Midgley picks up this theme in "The Religion of Evolution", built around the thesis that "Evolution is the creation-myth of our age". Midgley points out that scientists themselves are sometimes guilty of misuse of science to support world-views. Once they get beyond the technical aspects of their books, she alleges, they abandon all their careful intellectual standards and engage in the wildest speculation and prophesy. Durant points out that, ironically, such misuses of evolutionary science to legitimate world-views encourages the forces of religious anti-evolutionism.

Another central motif is that Darwinism and divinity have often supported each other. John Hedley Brooke, in a particularly strong study, shows that Darwin was not simply reacting against Paley's argument for God's existence based on the design in nature, but that Paley's thought provided some of the structure for Darwin's own analysis. Darwin's loss of faith in providential guidance, which Brooke looks at closely, was connected largely to his moral revulsion at a God who would permit so much suffering. Since widespread death and suffering was not a discovery original to the nineteenth century, it may be added, this moral objection (common among Darwin's contemporaries) may have had as much to do with the character of nineteenth-century upper-class Protestantism as with Darwin's technical work.

The more substantial positive relationship between Darwinism and divinity has been in the many accommodations that theologians have made between biological evolution and Christian faith. A. R. Peacocke, in another of the outstanding contributions, traces some of these. He also presents his own view. However the Universe started, it had at least the potential to produce human beings who can think about it. The postulate of God is necessary to make sense of the "intellectually coherent and explorable existence which science continually unveils". God is continually creating the world, relating to it as a composer relates to a performance of his music. The performance is both independent of him and intimately an expression of his creativity. God is an improviser of unsurpassed ingenuity, working out through evolutionary processes all sorts of potentialities in his universe.

The past century has produced many such reconciliations. Why then has the mythology prevailed that evolutionary theories and Christianity are usually in conflict? One of the many reasons becomes clearer from Jim (formerly James R.)

Moore's essay on the use of evolution by liberal Protestants in the late nineteenth century. Moore, always interesting, suggests that the great disaster for biological evolution was that it was so successfully adapted, especially by theologians, to supporting liberal culture. When liberal culture was later perceived to fail, the orthodox canons of its scientific basis understandably came under attack from everyone from Marxists to fundamentalists.

A similar point is made in Eileen Barker's insightful essay on the fundamentalist attacks on modern biology. Barker also observes that science today is unrivalled as a standard for epistemology, but is distrusted as a source of metaphysics. Resentment of science is also related to "almost a religious mystique about the inner gnoses of modern science", a problem exacerbated by everyone's dependence on the authority of specialists. When, in addition, natural science is so widely used to support various claims about the world, it seems to many to be operating on a religious plane.

Reshaking the kaleidoscope turns up some odd relationships. In an apparent fit of fairness to all viewpoints, the editor includes an essay by an anthropologist and a sociologist that attempts, plausibly but simplistically, to explain religious teachings on sexuality in terms of their function for promoting human survival. The authors use modern data to explain some ancient religious teachings and their reductionist assumptions seem to exemplify tendencies that other authors deplore.

The "kaleidoscopic" organization of the book provides only glimpses of the subject and will not fully satisfy those looking for systematic analysis. Nonetheless, the volume succeeds in its apparent purpose. It points out, often with insight, that recent cultural exploration of the history of science reveals relationships between modern biology and religion that have many more sides than the old warfare metaphors allowed us to see. □

George M. Marsden is a Professor in the Department of History, Calvin College, Grand Rapids, Michigan 49506, USA.

New in paperback

- *Powers of Ten: About the Relative Size of Things in the Universe*, by Philip Morrison and Phyllis Morrison. Publisher is Scientific American Books, price is \$19.95. For review see *Nature* 306, 125 (1983).
- *Alexander Fleming: The Man and the Myth*, by Gwyn Macfarlane. Publisher is Oxford University Press, price is £4.95. For review see *Nature* 308, 804 (1984).
- *100 Billion Suns: The Birth, Life and Death of the Stars*, by Rudolf Kippenhahn. Publisher is Counterpoint (Unwin Paperbacks), price is £4.50. For review see *Nature* 305, 252 (1983).
- *Bones*, by P.D.F. Murray, published by Cambridge University Press in their *Science Classics* series. Price is £12.50, \$19.95.

Change in relations

Brian Mathew

The Families of the Monocotyledons: Structure, Evolution and Taxonomy. By R.M.T. Dahlgren, H.T. Clifford and P. F. Yeo. *Springer-Verlag*:1985. Pp.520. DM294, \$103.20.

IT HAS long been acknowledged by students of the monocotyledonous section of the flowering plant kingdom that the classification, especially at family level, is very unsatisfactory. One can pick several areas where obviously fairly unrelated groups have been lumped together into a heterogeneous mass and the family label stuck upon it; in particular the Liliaceae comes to mind where, in the traditional sense, every petaloid monocotyledon with six stamens and a superior ovary was deposited. True lilies thus shared the same family bed as asparagus, hyacinths and yucca, to mention but a few of the more well-known and very divergent groups.

Although the recognition of smaller, more homogeneous families is not something new (Hyacinthaceae for example dates from 1802), the present study makes use of a wide range of modern data to give convincing explanations for their acceptance. The proposed new classification is, to say the least, alarming at first sight since some genera which were formerly placed

together in one family are now in some instances separated at order level and given family status of their own. A notable "split" which will surprise many botanists is between the Order Asparagales and the Liliales; Asparagales contains not only *Asparagus* but also many of the former members of the Liliaceae and Amaryllidaceae (both in Liliales). The decision to recognize the Asparagales as a group is strongly influenced by the presence of the blackish carbon-based compound phytochrome in the seed epidermis; this however is not the sole basis for its recognition, and as one works through this extremely meaty tome it is clear that the authors have studied character states from many disciplines and used them with much discretion in forming their suggested classification.

There are excellent line drawings illus-

trating many of the recognized plant families and there are also some interesting phylogenetical diagrams. Keys to the families are provided, but although these are particularly helpful it is perhaps optimistic of the authors to think that they are practical enough to be useful "in the field" when, for example, one of the dichotomies in the key to the Asparagales includes "Chromosome complement strongly dimorphic" versus "Chromosomes more uniform in size".

There are still many ifs, buts and maybe's. But this study represents one of the major steps forward in the understanding of the present relationships and the evolutionary pathways in the monocotyledons. □

Brian Mathew is a Principal Scientific Officer in the Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK.

The light fantastic

Ann Wintle

Thermoluminescence of Solids. By S.W.S. McKeever. *Cambridge University Press*: 1985. Pp.376. £40, \$69.50.

OVER the past 30 years people from a variety of different disciplines have attempted to use thermoluminescence (TL) as a research tool. The basis of the method is the ability of a large number of insulators and semiconductors to produce a measurable light signal when they are heated after exposure to radiation. Although there had been many observations of TL in naturally occurring minerals before the Second World War, it was not until the 1950s that several possible applications were seriously considered. Since then the technique has been used, over-extended and occasionally misused. In this book McKeever has brought together for the first time accounts of all the various uses of TL.

Research into applications as diverse as measurement of radiation doses in the natural environment and in medical treatment, dating of pottery from archaeological sites and the characterization of meteorites is carried out by different groups of physicists; they hardly ever attend the same conferences, or even read the same journals, and yet their research has the same basis and they use the same analytical procedures. McKeever states that his book is intended as a step "towards the unification of the insular approaches by presenting the topic of thermoluminescence as a single subject". I think he succeeds.

The book is fluently written, well-produced, clearly laid out in nine chapters and includes over 1,000 references, most of them from the past ten years. Three of the first four chapters provide the historical and theoretical background on which

the applications described in the remaining chapters rely. Chapter 3 deals with the experimental determination of the basic parameters which characterize the shapes of the glow curves, and which are used to predict the stability of TL signals either for dosimetry measurements or dating. Chapter 6 provides a comprehensive review of TL phosphors and their use in environmental monitoring, personal dosimetry and medical applications. This includes discussion of the responses to different types of radiation. The subsequent chapter covers the use of naturally occurring dosimeters for dating heated archaeological materials and the recent extension to dating unheated sedimentary deposits. This empirical approach contrasts with studies on alkali halides and quartz for which TL is just one of the methods for characterizing the defect structure.

The complexity of the TL properties of geological materials is exemplified in a further chapter which deals with meteoritic and lunar samples as well as terrestrial ones. In particular, stony meteorites have been the subject of a great deal of work since they are potentially the source of a considerable amount of information, and McKeever critically reviews the relevant literature. It is clear that for all natural matter the experimental TL worker is at the mercy of nature; however, detailed studies of the basic TL properties of the individual minerals can provide useful information on the overall thermal, optical and radiation histories of such materials.

The technical nature of the book does not detract from its readability and it provides an excellent, critical introduction to the subject for graduate students. With its extensive bibliography, it will also be a helpful work of reference for anyone with more than a passing interest in the diverse fields of applied thermoluminescence. □

Ann Wintle is at the Godwin Laboratory, University of Cambridge, Free School Lane, Cambridge CB2 3RS, UK.

THE ELEMENTS OF GRAPHING DATA

William S. Cleveland

Presents fundamental principles of graphical perception, describes principles that enhance graphical communication, and presents a wide variety of modern graphical methods useful for data analysis. Principles illustrated by many fascinating sets of data drawn from diverse areas of science and technology. Whether you use graphs for data analysis—or wish to communicate data to others—you will find that the graphical methods and principles are powerful tools that will increase your effectiveness in analyzing and reporting data.

Contents: 1. Introduction 2. Principles of Graph Construction 3. Graphical Methods 4. Graphical Perception

July 1985. 336 pp. cloth \$27.95, paper \$18.95.

Add sales tax for CA, KY, MA, MI, NC, NJ, NY, WA

W Wadsworth
Box N, 555 Abrego St.
Monterey, CA 93940

Law of the Sea threatens research

from Robert E. Bowen

Details of regulations evolving within the UN convention suggest that the freedom of marine research will be compromised.

A RECENT development at the Preparatory Commission of the International Seabed Authority (PrepCom) may have alarming implications for the conduct of marine scientific research in international waters. At its spring meeting in Jamaica, PrepCom began discussing regulations for mineral prospecting of the international seabed, called the "Area". The proposed definition of such activities is at present so broad and vague that virtually any marine research in the Area could be construed as falling under regulation by the International Seabed Authority (ISA).

As a result, further complications may be in store for marine science, not only for researchers from signatory states but for all those who seek to make studies in international waters. PrepCom is negotiating the operating rules and regulations for the International Seabed Authority, specified by the UN Convention on the Law of the Sea and which will be responsible, when the convention enters into force, for the conduct of mining activities in marine areas beyond national jurisdiction.

Many had thought that with the signing of the Law of the Sea convention, in 1982, a quarter of a century of international debate had ended. The convention aimed to define the direction of national and international marine law for the foreseeable future. For the most part, such expectations are still tenable. The Third United Nations Conference on the Law of the Sea (UNCLOS III) was able to reach agreement on the breadth of the territorial sea, the establishment of the Exclusive Economic Zone and principles regulating most fisheries stocks and the use of the sea for transport. The marine scientific community was assured that rules and procedures regulating the conduct of marine science had also been generally agreed upon; within areas of national jurisdiction, scientists must secure consent from the coastal state to carry out research, but outside national jurisdiction no such consent was required. However, the recent proposals for regulating prospecting seem to suggest that these agreements may be open for renegotiation.

At the latest meeting of the Preparatory Commission (11 March to 4 April 1985 at Kingston, Jamaica) certain draft articles began to take shape that could significantly affect the conduct of marine science outside national jurisdictions. These developments concern the powers of the ISA to regulate "prospecting activities" for all resources in the international ocean.

The difficulty lies with the definition of the term. According to the current working paper on the subject, "prospecting means the taking of geophysical, geochemical, oceanographic or atmospheric measurements, the collecting of rock, sediment and mineral samples from the surficial layers of the sea-bed, and establishing maps of data and sample locations, provided that such activities do not significantly alter the surface or subsurface of the sea-bed or significantly affect the environment or remove an appreciable quantity of material, *for the purpose of evaluating the exploitability of the resources of a specific area*" (my italics).

From such general wording, it is clear that much of marine science — most of marine geology, geophysics and geochemistry, for example — may be characterized as prospecting. Rather than a clear statement of intent, the draft article states that if the information derived from a scientific mission is used to evaluate the economic significance of mineralizations, then the activity could be viewed as prospecting and, thereby, subject to regulation by the ISA. Prospecting regulations would include such obligations as notification to and approval from the authority and the submission of annual reports. As presently conceived, these reports would include:

- The status of prospecting activities, including the amount of material recovered for analysis.
- Information on the degree to which the mission complies with ISA regulations.
- Any information obtained during prospecting relating to the protection of the environment.
- Reports on any training carried out as part of the mission.
- Observations on any activities affecting safety at sea.

The draft provisions also include a recognition that the processing of notifications will entail administrative costs on the part of the ISA, and that a fee schedule may be constructed.

The major problem lies in the inability of PrepCom to distinguish clearly between scientific research and inquiry directed exclusively at the development of information for future minerals exploitation. In a sense, this is a formal attempt to define the difference between basic and applied marine science.

The delegates to UNCLOS III had faced a similar dilemma. That debate proved to be so difficult that it was decided

not in scientific, but in geographical terms. It was resolved that activities within the bounds of national jurisdiction would ordinarily be regarded as applied science, but that it would be for the coastal state to allow exceptions within its marine science consent regime. It was also agreed that, outside national jurisdiction, no country or international organization had regulatory competence, so that scientific research would remain free. With the new developments, however, this old debate is rejoined. It is likely, but not certain, that PrepCom will draft formal recommendations to the ISA on general prospecting rules. Even if the United States and certain other countries (such as the United Kingdom and West Germany) remain outside the convention, the effect of these provisions, if accepted formally by the ISA, could be substantial. At the least, marine science conducted jointly with scientists from signatory countries (which include Canada, Mexico, France and Japan) could be affected or curtailed. This situation can be regarded as one cost of the refusal of critical deep-water-science states to participate actively in PrepCom bargaining. The United States has refused even to observe the proceedings, although permitted to attend the Final Act of the convention. Consequently, the interests of the international marine science community have not been well represented.

There may not be much time to put things right. At least three separate surveys of the international treaty ratification process (including those carried out by the UN secretariat and the government of Australia) suggest that the requisite number of ratifications to the convention will be deposited by 1989, which means that the treaty will probably be in force by 1990. The United States and others may be losing an important opportunity to influence the development of an international regulatory structure that could directly affect the conduct of marine science by their nationals. Without a more conscious effort by deep-water-science states, the international science community appears to have to depend on informal, but nonetheless important, avenues of approach. □

Robert E. Bowen is a visiting investigator at the Marine Policy and Ocean Management Center, Woods Hole Oceanographic Institution, Woods Hole Massachusetts 02543, USA, and a member of the Environmental Sciences Program at the University of Massachusetts at Boston Harbor Campus, Boston, Massachusetts 02125, USA.

Molecular mechanisms of receptor desensitization using the β -adrenergic receptor-coupled adenylate cyclase system as a model

David R. Sibley & Robert J. Lefkowitz

Howard Hughes Medical Institute, Departments of Medicine and Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

Desensitization, the tendency of biological responses to wane over time despite the continuous presence of a stimulus of constant intensity, is observed in organisms as diverse as bacteria and mammals. Recently, new insights into the molecular mechanisms underlying these phenomena have emerged from the study of the receptors coupled to the ubiquitous second messenger-generating system adenylate cyclase. These mechanisms involve sequestration or down-regulation of the receptors from the cell surface as well as functionally significant covalent modifications of the receptors and/or guanine nucleotide regulatory proteins.

DESENSITIZATION or adaptation is widespread in biological regulation, being found in, for example, chemotaxis of bacteria or mammalian polymorphonuclear leukocytes, neurotransmission by various neurotransmitters at synapses, stimulation of diverse physiological processes in eukaryotes by many drugs and sensory perception. Desensitization significantly limits the clinical use of many pharmacological agents.

Common to most systems that display desensitization is the existence of receptors that mediate the effects evoked by the specific stimuli. As such receptors constitute the first point of interaction of biologically active stimuli with cells, it is reasonable to suppose that regulation of receptor function constitutes the basis for some forms of desensitization.

Among the receptor-effector systems that mediate the effects of many hormones and drugs in man and other animals few are more important than the adenylate cyclase system, which synthesizes the ubiquitous second messenger cyclic AMP. Stimulation of the enzyme by catecholamines such as adrenaline and noradrenaline is mediated by a specific receptor termed the β -adrenergic receptor. Persistent stimulation of the system by catecholamines or by synthetic analogues leads to rapid desensitization of the cAMP response with consequent blunting of physiological responses, for example, the bronchodilator effect of adrenaline in the treatment of asthma or the inotropic effect of adrenergic agents in the treatment of congestive heart failure.

Recent progress in elucidating the molecular properties and functions of these receptors has shed new light on the mechanisms by which regulation of the receptors leads to desensitization. Moreover, the mechanisms uncovered may be generally operative in mediating the desensitization response in diverse biological systems. Here we review recent advances in understanding the molecular basis of desensitization, focusing on the receptor-coupled adenylate cyclase effector system (see refs 1-4 for further reviews).

Hormone-sensitive adenylate cyclase

Before considering desensitization of adenylate cyclase, we review briefly the structural components of this system. Hormone receptor-coupled adenylate cyclase seems to be composed of at least three proteins: (1) receptors, for example, the β -adrenergic receptors for catecholamines; (2) guanine nucleotide-binding regulatory proteins (N_s and N_i); and (3) the enzyme catalytic unit (C) (Fig. 1). β -Adrenergic receptors have been purified recently from various sources and seem to reside on single polypeptides of relative molecular mass (M_r) $\sim 60,000$ - $65,000$

from mammalian tissues^{5,6}. In amphibian and avian erythrocytes the receptors consist of peptides of M_r 58,000 (ref. 7) and $\sim 40,000$ - $50,000$ (refs 8, 9), respectively. Guanine nucleotide-binding proteins are involved in both the stimulation (N_s) and inhibition (N_i) of adenylate cyclase activity. These proteins have also been purified and shown to exist as heterotrimers with subunit M_r s of 42,000 (α_s) or 41,000 (α_i), 35,000 (β_s or β_i) and $\sim 5,000$ (γ_s or γ_i)^{10,11}. Less progress has been made in the characterization of the catalytic unit although a recent report of purification has indicated an M_r of $\sim 150,000$ (ref. 12).

Recent information concerning the activation of adenylate cyclase suggests that agonist binding to the receptor initiates the interaction of this protein with N_s . The coupling of these two proteins is readily detectable by assessing high-affinity, guanine nucleotide-sensitive agonist binding to the receptor. The subsequent binding of guanine nucleotides to N_s is thought to result in the dissociation of N_s from the receptor and the dissociation of the α - (GTP-binding) from the β - and γ -subunits of N_s . The guanine nucleotide-liganded α_s then combines with free or inactive catalytic units, which leads to stimulation of adenylate cyclase activity. Activity is terminated by a GTPase activity present in α_s . The interaction of N_s with the catalytic unit can be assessed by measuring guanine nucleotide- or fluoride-ion (which interacts with N_s) stimulation of adenylate cyclase activity. (For more detailed discussion of this topic, see refs 11, 13 and 14.)

Patterns of desensitization

An important problem in any mechanistic discussion of desensitization is that of nomenclature. Hormone-induced desensitization has usually been divided into two general categories referred to as agonist-specific or homologous desensitization and agonist-nonspecific or heterologous desensitization. (Desensitization of cAMP-mediated cellular responses could occur through a reduction in the rate of cAMP synthesis, enhancement of cAMP degradation or egress of cAMP from the cell. Although the latter two events are known to occur in some systems, it seems as if the major regulating factor is the modulation of the rate of receptor-stimulated cAMP synthesis.) The term 'homologous' is used to designate that form of desensitization in which only the subsequent response to the desensitizing hormone is attenuated while the efficacy of other hormone activators is unimpaired. Conversely, heterologous desensitization indicates that incubation with one agonist attenuates the response to multiple, different agonists operating through distinct receptors. Moreover, in some instances the

pattern of unresponsiveness of adenylate cyclase in heterologous desensitization may be so broad as to include decreased sensitivity to activators that bypass the receptors, for example, fluoride ion or guanine nucleotides. Although the terms homologous and heterologous are essentially phenomenological, we will retain this distinction between the two forms of desensitization in the following discussion of their biochemical mechanisms.

Homologous desensitization

The status of the β -adrenergic receptor during desensitization was first investigated in frog erythrocyte membranes after the cells were incubated with isoproterenol under conditions that lead to a homologous form of desensitization¹⁵⁻¹⁸. There was a decrease in the number of assayable β -adrenergic receptors directly comparable to the decrease in hormone responsiveness¹⁵⁻²⁰. Moreover, the time course of loss of β -adrenergic receptors paralleled the time course of loss of isoproterenol-stimulated adenylate cyclase activity^{18,19}. In contrast, homologous desensitization in cultured mammalian cells exhibits more complex phenomena consisting of multiple sequential events^{2-4,21-26}. An initial step involves a rapid 'uncoupling' of the receptors from the other components of the adenylate cyclase system, which may involve sequestration of receptors from the cell surface (see below). This can be measured as a loss in hormone-sensitive adenylate cyclase activity and a loss in high-affinity, nucleotide-sensitive agonist binding^{23,24}. A subsequent event involves a down-regulation of receptor number exhibited by a decrease in the ability to detect β -adrenergic receptors with radiolabelled antagonists^{21,22,24,25}. These processes seem to occur in all normal cells that exhibit homologous desensitization, although their kinetics can vary considerably. Homologous desensitization is thus associated with alterations occurring at the receptor level of the adenylate cyclase system.

The availability of a variant line (*cyc*⁻) of the S49 lymphoma cell that lacks the nucleotide regulatory protein N_s and hence hormone-sensitive adenylate cyclase activity has provided an important tool for probing many aspects of receptor-adenylate cyclase function. Using this cell line, Green and Clark^{27,28} have shown that agonist occupation of the β -adrenergic receptor in the absence of receptor- N_s coupling or cAMP generation is sufficient to promote desensitization in these cells. As *cyc*⁻ cells lack a functional N_s protein, hormone responsiveness can only be tested after reconstituting native N_s from wild-type cells with *cyc*⁻ membranes. Catecholamine-stimulated adenylate cyclase activity was decreased when the *cyc*⁻ membranes were derived from cells that had been preincubated with agonist^{27,28}. As agonist desensitization of *cyc*⁻ occurred in the absence of a functional N_s protein or activation of adenylate cyclase, these data suggest that the defect in homologous desensitization is localized to the β -adrenergic receptor. This makes intuitive sense as any defect distal to the receptor protein (that is, nucleotide regulatory proteins or catalytic unit) would probably lead to a heterologous form of desensitization. Moreover, these data indicate that homologous desensitization is not cAMP-mediated.

Other results have also suggested that the N_s protein is unaltered during homologous forms of desensitization^{27,29}. Such studies have used reconstitution of N_s from desensitized cells into acceptor β -adrenergic receptor systems and have demonstrated no impairment of function.

Receptor sequestration

The fact that the locus of the defect in homologous desensitization is the β -adrenergic receptor and the observation that receptors seem to decrease in number during desensitization suggests that the receptors are sequestered or internalized away from the cell surface. Initial work by Chuang and Costa³⁰⁻³² indicated that, in frog erythrocytes, loss of β -adrenergic receptors from the plasma membrane fraction was associated with an increase in binding activity in the cell cytosol. Stadel *et al.*³³ further showed that the β -adrenergic receptors that were located in the cytosolic fraction after desensitization in frog erythrocytes were sedimentable at 158,000g, indicating that these receptors

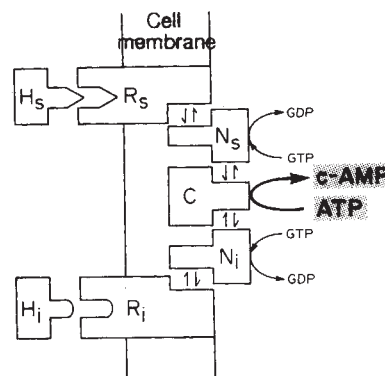


Fig. 1 Components of the hormone-responsive adenylate cyclase system. H, hormone; R, receptor; N, guanine nucleotide regulatory component; C, catalytic unit; s, stimulatory; i, inhibitory.

were located in light membrane particles or vesicles. These light membrane particles were devoid of adenylate cyclase activity and the guanine nucleotide regulatory protein N_s . These findings demonstrate that the receptors had been sequestered away from the cell surface into a membrane compartment that was lacking the other components of the adenylate cyclase system. Comparable findings have been obtained using the rat lung as an *in vivo* model of homologous desensitization³⁴.

Harden and collaborators³⁵⁻³⁷ have also provided results consistent with a sequestration of receptors during homologous desensitization in astrocytoma cells. Sucrose-density gradient centrifugation of control astrocytoma cell lysates resulted in a single peak of β -adrenergic receptor binding activity located with enzyme markers for plasma membranes including adenylate cyclase activity. In contrast, after desensitization to isoproterenol, the β -adrenergic receptors were present not only in the plasma membrane peak, but also in a peak sedimenting at light sucrose densities devoid of adenylate cyclase activity or other plasma membrane or cell surface markers. Similar findings have been reported by these authors and others using C6 glioma cells³⁸⁻⁴⁰.

Recently, additional evidence for a sequestration event has come from studies with the newly developed radioligand ³H-CGP-12177 (refs 41, 42). CGP-12177 is a relatively hydrophilic antagonist compared with the more hydrophobic antagonist radioligands used previously to study β -adrenergic receptors. Because of its relative hydrophilicity, ³H-CGP-12177 would be expected to label only the β -adrenergic receptors present on the cell surface when binding assays are performed on intact cells. Conversely, hydrophobic antagonists might be expected to penetrate into the cell and additionally label those receptors that are sequestered. Hertel *et al.*^{39,40} have shown, using C6 glioma cells, that isoproterenol treatment rapidly induces a loss of β -adrenergic receptors as assessed by binding of ³H-CGP-12177 to intact cells, which parallels the decrease in hormone-stimulated adenylate cyclase activity. However, the total number of receptors determined by binding of the hydrophobic antagonist radioligand ³H-dihydroalprenolol does not change.

An additional finding which may be related to the agonist-induced sequestration of β -adrenergic receptors is the observation that, as assessed by radioligand binding studies on intact cells, the affinity of the receptors for agonists declines with time during agonist exposure to intact cells⁴³⁻⁴⁶. Preincubation of cells with agonists converts about half the β -adrenergic receptors from a form exhibiting high affinity for the agonists to a form with lower apparent affinity. Antagonists do not have this effect. These results have been interpreted to indicate that the low-affinity form of the β -adrenergic receptor observed in such whole-cell experiments results from slow equilibration of the hydrophilic agonists with a population of β -adrenergic receptors that have been internalized away from the cell surface because of agonist exposure. Interestingly, this effect occurs in the S49 lymphoma *cyc*⁻ variant cells that lack a functional N_s pro-

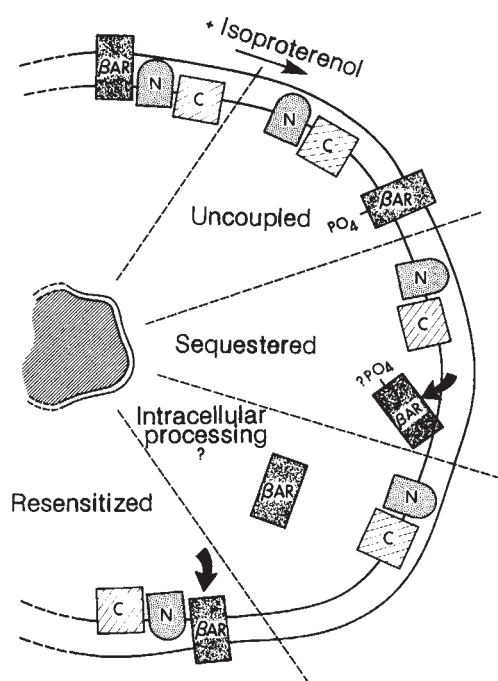


Fig. 2 Homologous desensitization of β -adrenergic receptor (BAR)-coupled adenylate cyclase.

tein^{47,48}. Apparently, the rapid change in affinity of the β -adrenergic receptor for agonists requires neither the interaction of the receptor with N_s nor the activation of adenylate cyclase and generation of cAMP. This is consistent with observations that showed that agonist-induced desensitization does not require β -adrenergic receptor- N_s coupling in S49 cells^{27,28}.

Receptor down-regulation

As well as an early desensitization/sequestration event, homologous desensitization in some cells seems to be associated with an actual disappearance or loss of receptors such that they cannot be detected with any radioligand in any cell fraction^{21,22,24,25}. This phenomenon has come to be referred to as 'down-regulation' of the receptors as opposed to the sequestration of receptors described above. Down-regulation occurs more slowly than the initial sequestration process in most cell types, and the exact fate of these β -adrenergic receptors is not known with certainty. In some cases it seems clear that the receptors are proteolytically degraded and new receptor synthesis is required to regenerate them^{38,49,50}. In others, the number of β -adrenergic receptors can return to control levels even in the absence of new protein synthesis^{50,51}. This suggests that these down-regulated receptors can regain their binding activity with time. Interestingly, in contrast to the sequestration event, β -adrenergic receptor- N_s coupling seems to be required for maximal down-regulation to occur. Thus, in S49 *cyc*⁻ and *unc* variant cells, which possess lesions in N_s , agonist-induced down-regulation of β -adrenergic receptors does not occur or is blunted despite the fact that these cells apparently exhibit the desensitization/sequestration process^{21,24,52}. This suggests that the down-regulation process involves an N_s -mediated targeting of the receptors through an as yet undefined mechanism.

Although β -adrenergic receptor- N_s coupling seems to be necessary for down-regulation, activation of adenylate cyclase and the generation of cAMP is not required. Experiments with HC-1 hepatoma cells that possess β -adrenergic receptor and N_s but not a functional catalytic unit demonstrated that isoproterenol pretreatment could produce down-regulation of hydrophobic antagonist binding²⁴. Additionally, the S49 variant *kin*⁻, which lacks cyclic AMP-dependent protein kinase, exhibits desensitization and down-regulation of β -adrenergic receptors in response to catecholamine exposure, indicating that this kinase is not involved in these processes^{21,52}.

Receptor uncoupling

Homologous desensitization thus seems to be composed of multiple related events serving to sequester or remove the receptor from its normal position in the plasma membrane. Although the receptor is eventually sequestered from N_s and C during the desensitization process, a central question relates to whether the receptor is functionally modified as well. Strulovici *et al.*⁵³ have tested the functional activity of the sequestered receptors that can be recovered in light membrane particles in frog erythrocytes. By fusing these receptors to a cell (*Xenopus laevis* erythrocyte) that possesses the adenylate cyclase components N_s and C but lacks β -adrenergic receptors, sequestered receptors were shown to be functionally active. Similar conclusions have been reached by Clark *et al.*⁵⁴ in characterizing the sequestered receptors found in S49 cell light membrane fractions. Strasser and Lefkowitz⁵⁵ have also found that treatment of desensitized S49 cells with the fusogen polyethylene glycol is sufficient to restore the catecholamine-stimulated enzyme activity to control levels. Such studies suggest that homologous desensitization results primarily from the physical sequestration of the receptors. However, results of such studies must be interpreted in the light of the caveat that the reconstitution and/or fusion procedures might reverse a functional alteration.

In contrast, Kassiss and Fishman⁵⁶, using membrane-membrane and membrane-cell fusion techniques, have demonstrated that catecholamine-induced homologous desensitization in various cultured cells does indeed result in a functional alteration of the β -adrenergic receptor. Perkins and colleagues^{36,37} have also presented evidence that very short (1–2 min) exposure of astrocytoma cells to isoproterenol leads to desensitization of hormone-sensitive adenylate cyclase, yet the β -adrenergic receptors are not sequestered at this time. Moreover, pre-treatment of astrocytoma cells with concanavalin A blocks the sequestration process in these cells, yet does not block the desensitization and uncoupling events which occur maximally. These data suggest that a functional change in the receptor occurs early in the desensitization process but that this modification might be reversed by the time sequestration is complete. It is thus not clear if physical sequestration alone can completely account for the receptor alterations observed in homologous desensitization.

Homologous desensitization mechanisms

The biochemical mechanisms involved in homologous desensitization are not yet known with certainty. Studies of homologous desensitization in cell-free systems have indicated a requirement for phosphorylating conditions^{57–59} and in at least one study⁵⁹, demonstrated that the desensitization could be reversed by phosphatase treatment. Recently, Sibley *et al.*⁶⁰ have shown directly that homologous desensitization of adenylate cyclase in frog erythrocytes is associated with phosphorylation of the β -adrenergic receptor. This phosphorylation is induced in a stereospecific fashion by isoproterenol and is blocked by the β -adrenergic antagonist propranolol. The phosphorylation state of the receptor stoichiometrically increases ~threefold on desensitization. Interestingly, prostaglandin E_1 (PGE₁) does not promote β -adrenergic receptor phosphorylation despite the fact that PGE₁ activates adenylate cyclase in these cells⁶⁰. This observation agrees well with the notion that homologous desensitization is not cAMP-mediated. The nature of the protein kinase that phosphorylates the β -adrenergic receptor during homologous desensitization is unknown.

There are at least two potential mechanisms by which receptor phosphorylation could contribute to the homologous desensitization. One possibility is that the phosphorylation induces or triggers the sequestration of the receptor from the cell surface. Another possibility is that the phosphorylation results in a functional modification in the receptor protein such that it is less efficacious in activating adenylate cyclase. If the phosphorylation promotes an early occurring functional change in the receptor then, as discussed above, the receptor phosphorylation

might be reversed on sequestration. Obviously, the relationship between receptor phosphorylation, sequestration and any functional modification will require the resolution of the desensitized receptors into plasma membrane and sequestered fractions and the determination of the phosphorylation states and functionality of each.

Another major question associated with homologous desensitization concerns the precise cellular location of the sequestered receptors. One hypothesis is that the receptors are internalized into intracellular vesicles via a clustering/coated pit mechanism as demonstrated for peptide hormone receptors⁶¹. The observation that on lysis of desensitized cells, the sequestered receptors seem to have been translocated into light membrane particles suggests that this is the case. However, Strader *et al.*⁶² have shown that, in frog erythrocytes, the production of these vesicular particles is dependent on the method of cell lysis. Using gentle methods of lysis, the sequestered β -adrenergic receptors were shown to be localized in the plasma membrane fraction, indicating that in the intact cell the receptors are never present in free-floating cytoplasmic vesicles but rather remain contiguous with the inner surface of the plasma membrane. A definitive localization of the receptors must await development of a means of visualizing the β -adrenergic receptors in intact cells either by electron microscope or by fluorescent techniques.

Heterologous desensitization

Heterologous forms of desensitization of adenylate cyclase occur in many tissues and cell types. As discussed above, this form of desensitization represents a broad pattern of refractoriness in which the response to multiple hormones and sometimes nonhormonal effectors is impaired. In contrast to homologous desensitization (which may be unimechanistic), heterologous desensitization certainly occurs by more than one mechanism. In many cell types, heterologous desensitization occurs in addition to the homologous form of desensitization, thus making its analysis difficult. In general, however, the heterologous response occurs with a slower onset than the homologous one, suggesting that heterologous desensitization represents an adaptive response to relatively prolonged stimulation.

Perkins and colleagues^{2,51,63} were among the first to investigate heterologous desensitization using clonal astrocytoma cells. Prolonged incubation with either β -adrenergic agonists or prostaglandins diminished the subsequent capacity of both hormones to elevate intracellular cAMP levels. Incubation of the cells with dibutyl cAMP produced refractoriness to both catecholamines and prostaglandins, suggesting that heterologous desensitization is cAMP-mediated. Brooker and coworkers have also shown that in C6-2B rat glioma cells a heterologous form of adenylate cyclase desensitization can be elicited with catecholamines and other agents that elevate intracellular cAMP, including cAMP analogues, cholera toxin, forskolin and phosphodiesterase inhibitors⁶⁴⁻⁶⁸. No loss of β -adrenergic receptor binding was found.

Alteration of N_s

Cultured fibroblasts provide an important model system for clarifying some mechanisms of heterologous desensitization. In these cells, catecholamines induce homologous desensitization whereas PGE₁ induces a heterologous form of desensitization. Kassir and Fishman⁶⁹ demonstrated that when the N_s protein from PGE₁-desensitized fibroblasts is reconstituted into S49 *cyc*⁻ membranes, it is less efficient in restoring adenylate cyclase activity compared with control or catecholamine-treated cells. These results suggest that at least one mechanism of the heterologous desensitization in fibroblast cells is an impaired functionality of the N_s protein. Heterologous desensitization induced by prostaglandin E in liver⁷⁰ and by human chorionic gonadotropin in ovaries⁷¹ also results in an impaired functionality of N_s as determined with S49 *cyc*⁻ membrane reconstitution. In contrast, Rich *et al.*⁷² have reported that glucagon-induced heterologous desensitization in MDCK cells was not associated

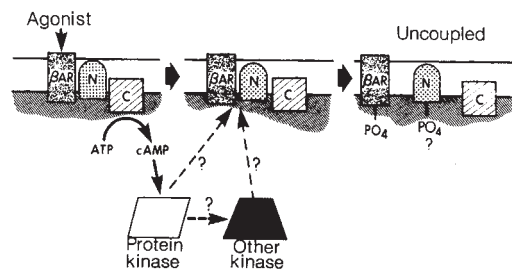


Fig. 3 Heterologous mechanisms of adenylate cyclase desensitization.

with N_s alterations but instead involved increases in the apparent levels of the inhibitory guanine nucleotide regulatory protein, N_i . This novel finding suggests that alterations in the N_s/N_i stoichiometry are another mechanism by which heterologous desensitization can be achieved.

Additional evidence for a functional alteration in N_s in heterologous desensitization has come from studies with avian erythrocytes. Incubation of turkey, pigeon and duck erythrocytes with catecholamines leads to pronounced desensitization of catecholamine-stimulated adenylate cyclase activity in membrane preparations, with no loss of β -adrenergic receptors⁷³⁻⁷⁵. Although only β -adrenergic receptors are coupled to adenylate cyclase in these cells, there are also significant decrements in guanine nucleotide- and fluoride ion-stimulated enzyme activities, indicating a heterologous form of desensitization. Hudson and Johnson initially found that desensitization of pigeon erythrocytes did not alter the ability of solubilized N_s to reconstitute adenylate cyclase activity in S49 *cyc*⁻ membranes⁷⁵. They did find, however, that desensitization resulted in a diminished ability of agonists and/or guanine nucleotides to induce a conformational change in the N_s protein. Different results were obtained by Briggs *et al.*⁷⁶, who found that when the solubilized N_s protein of turkey erythrocytes was carefully quantitated by labelling with ³²P-NAD⁺ and cholera toxin, desensitization resulted in a significant reduction of the ability of N_s to reconstitute enzyme activity in S49 *cyc*⁻ membranes. Thus, as with the fibroblast cells, heterologous desensitization in avian erythrocytes is associated with an impairment in N_s function.

Receptor alterations

Although desensitization of avian erythrocytes results in impaired N_s functionality, this does not exclude other potential lesions in the adenylate cyclase system. In fact, the observation that catecholamine-stimulated adenylate cyclase activity is desensitized by ~50% whereas the fluoride ion and guanine nucleotide activities are reduced by only 10-20% after agonist-induced desensitization is indicative of other processes that occur as well. Stadel *et al.*⁷⁷ first provided evidence for this possibility by demonstrating that the β -adrenergic receptor peptides, detected by photoaffinity labelling and SDS-polyacrylamide gel electrophoresis, were of larger apparent M_r after desensitization of turkey erythrocytes, indicating that the desensitization process coincided with a stable structural modification of the receptor.

Additional evidence for a covalent/functional modification of the receptor protein has been obtained by partially purifying the β -adrenergic receptors from desensitized turkey erythrocytes, reconstituting them into phospholipid vesicles and fusing them with *X. laevis* erythrocyte acceptor cells that possess N_s and C but no β -adrenergic receptors⁷⁸. The β -adrenergic receptors from desensitized cells were less efficient than those from control cells in reconstituting catecholamine responsiveness of the *X. laevis* adenylate cyclase activity.

β -Adrenergic receptor phosphorylation

Using ³²P incorporation it was shown that the β -adrenergic receptor undergoes phosphorylation during the desensitization process in turkey erythrocytes⁷⁹; Sibley *et al.*⁸⁰ have investigated

this process in detail. The β -adrenergic receptor is stoichiometrically phosphorylated under 'basal' conditions containing 0.7–1.0 mol phosphate per mol receptor, with this stoichiometry increasing to 2–3 mol per mol on maximal catecholamine-induced desensitization. Moreover, the phosphate/receptor stoichiometry is tightly correlated with the degree of desensitization as tested under many experimental conditions⁸⁰. Interestingly, incubation of the cells with cAMP analogues promotes only partial phosphorylation of the β -adrenergic receptor, which is correlated with the partial desensitization of adenylate cyclase that these compounds evoke⁸⁰. These data thus suggest that in this system, heterologous desensitization is closely associated with phosphorylation of the β -adrenergic receptor.

Additional evidence for a role of phosphorylation in avian erythrocyte desensitization has come from experiments using a cell-free model system. Nambi *et al.*^{81,82} demonstrated catecholamine-induced desensitization in turkey erythrocyte lysates and showed that this process required added ATP, Mg^{2+} and factor(s) present in the cytoplasmic fraction of the cell. If [γ -³²P]ATP was also used, isoproterenol could be shown to induce phosphorylation of the β -adrenergic receptor in a stereospecific fashion, similar to that observed using intact cells.

Recently, Benovic *et al.*⁸³ have examined a cAMP-mediated phosphorylation of mammalian lung β -adrenergic receptor that is similar to the process exhibited by avian erythrocytes. Using pure lung receptor and pure cAMP-dependent protein kinase, it was shown that isoproterenol could enhance the rate of receptor phosphorylation by ~twofold. Reconstitution of the phosphorylated receptor with the N_s protein demonstrated diminished agonist-promoted receptor-mediated stimulation of GTPase activity compared with controls.

An interesting question concerning avian erythrocyte desensitization is why cAMP analogues produce only partial effects. One intriguing hypothesis, suggested by the studies of Benovic *et al.*⁸³, is that to obtain maximal phosphorylation of the receptor, agonist occupancy must occur. Another possibility is that the receptor phosphorylation is not completely mediated by cAMP and that other protein kinase systems may also phosphorylate the β -adrenergic receptor. In this regard, we^{82,84} and others⁸⁵ have shown recently that tumour-promoting phorbol diesters, compounds that potentially activate protein kinase C, can stimulate β -adrenergic receptor phosphorylation and induce adenylate cyclase desensitization in avian erythrocytes. In duck erythrocytes, the phorbol diester-promoted β -adrenergic receptor phosphorylation was shown to be nonadditive with that produced by catecholamines, suggesting a common mechanism or pathway of action⁸⁴.

Heterologous desensitization mechanisms

Heterologous desensitization in avian erythrocytes is thus associated with modifications in the adenylate cyclase system at both the level of the receptor and the guanine nucleotide regulatory proteins. It should be pointed out that structural and/or functional alterations of the enzyme catalytic unit cannot be excluded although there is no evidence for this at present.

It seems reasonable to propose that the mechanisms of heterologous desensitization elucidated in erythrocytes are not unique to this cell type. Thus, one major means of achieving heterologous desensitization would be modification(s) in the N_s or N_i proteins as discussed above. The nature of these modifications is not yet known, although an attractive hypothesis is that a phosphorylation event is involved. Although, in some systems, cAMP seems to induce the N_s/N_i modification^{74,86}, in others it seems not to be involved^{72,87}.

Another major pathway of heterologous desensitization is modification of the receptor proteins. In this case the effect is not receptor sequestration or down-regulation as in homologous desensitization but a functional uncoupling of the receptor from adenylate cyclase. Evidence from avian erythrocytes indicates that phosphorylation is the probable mechanism for producing the functional modification. Moreover, because of the heterologous nature of the desensitization it is reasonable to

suppose that all the receptors coupled to adenylate cyclase would probably be modified in this way, that is, phosphorylated.

It is likely that differing extents of receptor and N_s/N_i modifications occur in different cell types and under different conditions. In some cells that exhibit heterologous desensitization the modification in N_s or N_i might predominate. In other cells, including the erythrocytes, modification of the receptor protein may be the most relevant mechanism. Evidence for the latter possibility has come from several systems that exhibit heterologous desensitization in the absence of any alterations in guanine nucleotide- or fluoride ion-stimulated enzyme activities^{72,88,89}.

Other examples of desensitization

As noted above, diverse biological phenomena are characterized by adaptive processes analogous to the desensitization observed in β -adrenergic receptor-coupled adenylate cyclase systems. Not surprisingly, many of the other adenylate cyclase-coupled receptors such as those for glucagon, vasopressin and parathyroid hormone display desensitization phenomena qualitatively similar to those described above. Thus far, little information is available concerning the molecular mechanisms involved. It will be of interest to see if these are identical to those uncovered for the β -adrenergic receptors.

α_1 -Adrenergic receptors are coupled to processes that lead to the turnover of phosphatidylinositol biphosphate and the generation of two potential second messengers: diacylglycerol, which activates protein kinase C, and inositol triphosphate, which mobilizes Ca^{2+} . Recently it has been demonstrated that when protein kinase C is activated by phorbol diesters in a line of cultured smooth muscle cells, the α_1 receptors are stoichiometrically phosphorylated and functionally uncoupled⁹⁰.

A fascinating analogy may be drawn with a potential mechanism of light adaptation recently elucidated for the rhodopsin-coupled effector system in the rod outer segments of the retina. As reviewed in detail elsewhere, this system is remarkably analogous to hormone responsive adenylate cyclases⁹¹. The analogous components are a photon of light instead of hormone, rhodopsin instead of receptor, a GTP-binding and hydrolysing protein termed transducin instead of N_s , and cGMP phosphodiesterase, which controls retinal concentrations of cGMP, instead of the effector adenylate cyclase enzyme. The exact way in which this system functions to transduce a light signal to a visual impulse is unclear. However, the functional and structural analogies of the system with the adenylate cyclase are striking. Thus, interaction of light with the 'receptor' rhodopsin precipitates a series of conformational changes in rhodopsin that ultimately leads to interactions with the coupling protein transducin. The latter protein then stimulates the activity of the phosphodiesterase. A fourth protein component of this system, the enzyme rhodopsin kinase, is capable of specifically phosphorylating rhodopsin and thereby inhibiting its ability to interact with and stimulate transducin^{92–94}. This phosphorylation reaction seems to require the bleached form of rhodopsin as its substrate, that is, the form that has interacted with the 'agonist' light. Thus, this system is desensitized by an agonist-promoted phosphorylation reaction that uncouples the receptor from its GTP-binding regulatory protein. This mechanism is strikingly parallel to that described above for desensitization of the β -adrenergic receptor by phosphorylation. Moreover, it raises the interesting possibility that, by analogy with the rhodopsin system, there might be an unidentified specific β -adrenergic receptor kinase involved in mediating receptor phosphorylation and desensitization.

A further example of how covalent modification of a receptor protein can lead to its desensitization or adaptation is found in bacterial chemotaxis receptors. These bacterial membrane proteins are able to 'sense' gradients of certain nutrients such as galactose, ribose or aspartate and influence the mobility of the cells so that they swim in the direction of the gradient. However, even as the intensity of the chemotactic stimulus increases, the

'signal' generated by the receptors tends to remain constant. This desensitization reaction seems to be mediated by a specific methylation of the receptor protein⁹⁵⁻⁹⁷. Again, by analogy with both the rhodopsin system and the β -adrenergic receptor-coupled adenylate cyclase system, the presence of an 'agonist' or effector on the receptors induces conformational changes that facilitate the subsequent covalent modification leading to the adaptation. Thus, even at this very primitive level of evolution, regulation of receptor protein function by covalent modification can be observed, thus presaging the existence of analogous mechanisms in vertebrates by hundreds of millions of years.

Summary

Multiple mechanisms seem to be involved in regulating the responsiveness of hormone receptor-coupled adenylate cyclase systems. These mechanisms at least involve the receptors and nucleotide regulatory proteins. With the recent development of methods for purifying the catalytic unit of the enzyme it will be possible to assess whether it is also a locus for such regulatory phenomena.

At least two major pathways of receptor regulation have been uncovered. Homologous desensitization (Fig. 2) involves the uncoupling and translocation of the receptors out of their normal plasma membrane environment. This process sequesters the receptors away from their effector, the regulatory and catalytic components of adenylate cyclase. The site of receptor sequestration is unclear and might lie within the plasma membrane or

within the cell. The sequestered receptors can recycle to the cell surface or become down-regulated, perhaps being destroyed within the cell. Phosphorylation of the receptors through a non-cAMP-mediated process has been shown to be associated with homologous desensitization. This phosphorylation event may serve either to uncouple functionally the receptors or to trigger their sequestration from the cell surface or both.

In heterologous desensitization (Fig. 3), receptor function is regulated by phosphorylation in the absence of receptor sequestration or down-regulation. This covalent modification serves to functionally uncouple the receptors, that is, to impair their interactions with the guanine nucleotide regulatory proteins. Several protein kinases seem to be capable of promoting phosphorylation of the receptors including the cAMP-dependent kinase and protein kinase C. In addition to the receptor modification, heterologous desensitization seems to be associated with functional modifications (phosphorylation?) at the level of the nucleotide regulatory proteins (N_s and N_i) (Fig. 3). These mechanisms are analogous to mechanisms that have evolved to regulate such disparate receptors as the light receptor rhodopsin and bacterial chemotaxis receptors by covalent modification. Further studies of the mechanisms of desensitization of adenylate cyclase-coupled receptors are thus likely to help elucidate modes of regulation of a wide variety of receptor-coupled functions in diverse types of cells.

We thank Lynn Tilley for secretarial assistance. D.R.S. is the recipient of NIH Postdoctoral Fellowship HL06631.

1. Lefkowitz, R. J., Wessels, M. R. & Stadel, J. M. *Curr. Top. Cell. Regul.* **17**, 205-230 (1980).
2. Perkins, J. P. *Curr. Top. Membranes Transp.* **18**, 85-108 (1983).
3. Harden, T. K. *Pharmac. Rev.* **35**, 5-32 (1983).
4. Hertel, C. & Perkins, J. P. *Molec. cell. Endocr.* **37**, 245-256 (1984).
5. Benovic, J. L., Shorr, R. G. L., Caron, M. G. & Lefkowitz, R. J. *Biochemistry* **23**, 4510-4518 (1984).
6. Cubero, A. & Malbon, C. C. *J. biol. Chem.* **259**, 1344-1350 (1984).
7. Shorr, R. G. L., Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **256**, 5820-5826 (1981).
8. Shorr, R. G. L., Strohsacker, M. W., Lavin, T. N., Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **257**, 12341-12350 (1982).
9. Sibley, D. R., Peters, J. R., Nambi, P., Caron, M. G. & Lefkowitz, R. J. *Biochem. biophys. Res. Commun.* **119**, 458-464 (1984).
10. Codina, J. et al. *Adv. Cyclic Nucleotide Res.* **17**, 111-126 (1984).
11. Gilman, A. G. *Cell* **36**, 577-579 (1984).
12. Pfeuffer, E., Dreher, R. M., Metzger, H. & Pfeuffer, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3086-3090 (1985).
13. Stadel, J. M. & Lefkowitz, R. J. *Curr. Top. Membranes Transp.* **18**, 45-66 (1983).
14. Schramm, M. & Selinger, Z. *Science* **225**, 1350-1356 (1984).
15. Mukherjee, C., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1945-1949 (1975).
16. Mukherjee, C., Caron, M. G. & Lefkowitz, R. J. *Endocrinology* **99**, 347-357 (1976).
17. Mickey, J., Tate, R. & Lefkowitz, R. J. *J. biol. Chem.* **250**, 5727-5729 (1975).
18. Mickey, J. V., Tate, T., Mullikin, D. & Lefkowitz, R. J. *Molec. Pharmac.* **12**, 409-419 (1976).
19. Wessels, M. R., Mullikin, D. & Lefkowitz, R. J. *J. biol. Chem.* **253**, 3371-3373 (1978).
20. Wessels, M. R., Mullikin, D. & Lefkowitz, R. J. *Molec. Pharmac.* **16**, 10-20 (1979).
21. Shear, M., Insel, P. A., Melmon, K. L. & Coffino, P. *J. biol. Chem.* **251**, 7572-7576 (1976).
22. Su, Y.-F., Harden, T. K. & Perkins, J. P. *J. biol. Chem.* **254**, 38-41 (1979).
23. Harden, T. K., Su, Y.-F. & Perkins, J. P. *J. Cyclic Nucleotide Res.* **5**, 99-106 (1979).
24. Su, Y.-F., Harden, T. K. & Perkins, J. P. *J. biol. Chem.* **255**, 7410-7419 (1980).
25. Homburger, V. et al. *J. biol. Chem.* **255**, 10436-10444 (1980).
26. Fishman, P. H., Mallorga, P. & Tallman, J. F. *Molec. Pharmac.* **20**, 310-318 (1981).
27. Green, D. A. & Clark, R. B. *J. biol. Chem.* **256**, 2105-2108 (1981).
28. Green, D. A., Friedman, J. & Clark, R. B. *J. Cyclic Nucleotide Res.* **7**, 161-172 (1981).
29. Iyengar, R., Bhat, M. K., Riser, M. E. & Birnbaumer, L. *J. biol. Chem.* **256**, 4810-4815 (1981).
30. Chuang, D.-M. & Costa, E. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3024-3028 (1979).
31. Chuang, D.-M., Kinnier, W. J., Farber, L. & Costa, E. *Molec. Pharmac.* **18**, 348-355 (1980).
32. Chuang, D.-M. & Costa, E. *Neurochem. Res.* **4**, 777-793 (1979).
33. Stadel, J. M. et al. *J. biol. Chem.* **258**, 3032-3038 (1983).
34. Strasser, R. H., Stiles, G. L. & Lefkowitz, R. J. *Endocrinology* **115**, 1392-1400 (1984).
35. Harden, T. K., Cotton, C. U., Waldo, G. L., Lutton, J. K. & Perkins, J. P. *Science* **210**, 441-443 (1980).
36. Waldo, G. L., Northup, J. K., Perkins, J. P. & Harden, T. K. *J. biol. Chem.* **258**, 13900-13908 (1983).
37. Toews, M. L., Waldo, G. L., Harden, T. K. & Perkins, J. P. *J. biol. Chem.* **259**, 11844-11850 (1984).
38. Frederich, R. C. Jr, Waldo, G. L., Harden, T. K. & Perkins, J. P. *J. Cyclic Nucleotide Prot. Phosphor. Res.* **9**, 103-118 (1983).
39. Hertel, C., Staehelin, M. & Perkins, J. P. *J. Cyclic Nucleotide Prot. Phosphor. Res.* **9**, 119-128 (1983).
40. Hertel, C., Muller, P., Portenier, M. & Staehelin, M. *Biochem. J.* **216**, 669-674 (1983).
41. Staehelin, M., Simons, P., Jaegg, K. & Wigger, N. *J. biol. Chem.* **258**, 3496-3502 (1983).
42. Staehelin, M. & Hertel, C. *J. receptor Res.* **3**, 35-43 (1983).
43. Pittman, R. N. & Molinoff, P. B. *J. Cyclic Nucleotide Res.* **6**, 421-435 (1980).
44. Pittman, R. N. & Molinoff, P. B. *Molec. Pharmac.* **24**, 398-408 (1983).
45. Toews, M. L., Harden, T. K. & Perkins, J. P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3553-3557 (1983).
46. Toews, M. L. & Perkins, J. P. *J. biol. Chem.* **259**, 2227-2235 (1984).
47. Insel, P. A., Mahan, L. C., Motulsky, H. J., Stoolman, L. M. & Koachman, A. M. *J. biol. Chem.* **258**, 13597-13605 (1983).
48. Hoyer, D., Reynolds, E. E. & Molinoff, P. B. *Molec. Pharmac.* **25**, 209-218 (1984).
49. Morishima, I., Thompson, W. J., Robison, G. A. & Strada, S. J. *Molec. Pharmac.* **18**, 370-378 (1980).
50. Doss, R. C., Perkins, J. P. & Harden, T. K. *J. biol. Chem.* **256**, 12281-12286 (1981).
51. Su, Y.-F., Cubeddu, L. X. & Perkins, J. P. *J. Cyclic Nucleotide Res.* **2**, 257-270 (1976).
52. Mahan, L. C., Koachman, A. M. & Insel, P. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 129-133 (1985).
53. Strulovici, B., Stadel, J. M. & Lefkowitz, R. J. *J. biol. Chem.* **258**, 6410-6414 (1983).
54. Clark, R. B., Friedman, J., Prashad, N. & Ruoho, A. E. *J. Cyclic Nucleotide Prot. Phosphor. Res.* **10**, 97-119 (1985).
55. Strasser, R. H. & Lefkowitz, R. J. *J. biol. Chem.* **260**, 4561-4564 (1985).
56. Kassir, S. & Fishman, P. H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6686-6690 (1984).
57. Anderson, W. B. & Jaworski, C. J. *J. biol. Chem.* **254**, 4596-4601 (1979).
58. Salomon, Y., Ezra, E. & Amir-Zaltsman, Y. *Adv. Cyclic Nucleotide Res.* **14**, 101-109 (1981).
59. Hunzicker-Dunn, M., Derda, D., Jungmann, R. A. & Birnbaumer, L. *Endocrinology* **104**, 1785-1793 (1979).
60. Sibley, D. R., Strasser, R. H., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **260**, 3883-3886 (1985).
61. Pastan, I. H. & Willingham, M. C. A. *Rev. Physiol.* **43**, 239-250 (1981).
62. Strader, C. D., Sibley, D. R. & Lefkowitz, R. J. *Life Sci.* **35**, 1601-1610 (1984).
63. Johnson, G. L., Wolfe, B. B., Harden, T. K., Molinoff, P. B. & Perkins, J. P. *J. biol. Chem.* **253**, 1472-1480 (1978).
64. de Vellis, J. & Brooker, G. *Science* **186**, 1221-1223 (1974).
65. Terasaki, W. L. et al. *Adv. Cyclic Nucleotide Res.* **9**, 33-52 (1978).
66. Nichols, G. A. & Brooker, G. *J. Cyclic Nucleotide Res.* **5**, 435-447 (1979).
67. Nichols, G. A. & Brooker, G. *J. biol. Chem.* **255**, 23-26 (1980).
68. Barovsky, K., Pedone, C. & Brooker, G. *J. Cyclic Nucleotide Prot. Phosphor. Res.* **9**, 181-189 (1983).
69. Kassir, S. & Fishman, P. H. *J. biol. Chem.* **257**, 5312-5318 (1982).
70. Garrity, M. J., Andreasen, T. J., Storm, D. R. & Robertson, R. P. *J. biol. Chem.* **258**, 8692-8697 (1983).
71. Kirchik, H. J., Iyengar, R. & Birnbaumer, L. *Endocrinology* **113**, 1638-1646 (1983).
72. Rich, K. A. et al. *J. biol. Chem.* **259**, 7893-7901 (1984).
73. Hoffman, B. B., Mullikin-Kilpatrick, D. & Lefkowitz, R. J. *J. Cyclic Nucleotide Res.* **5**, 355-366 (1979).
74. Simpson, I. A. & Pfeuffer, T. *Eur. J. Biochem.* **111**, 111-116 (1980).
75. Hudson, T. H. & Johnson, G. L. *Molec. Pharmac.* **20**, 694-703 (1981).
76. Briggs, M. M., Stadel, J. M., Iyengar, R. & Lefkowitz, R. J. *Archs Biochem. Biophys.* **224**, 142-151 (1983).
77. Stadel, J. M. et al. *J. biol. Chem.* **257**, 9242-9245 (1982).
78. Strulovici, B., Cerione, R. A., Kilpatrick, B. F., Caron, M. G. & Lefkowitz, R. J. *Science* **225**, 837-840 (1984).
79. Stadel, J. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3173-3177 (1983).
80. Sibley, D. R., Peters, J. R., Nambi, P., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **259**, 9742-9749 (1984).
81. Nambi, P. et al. *J. biol. Chem.* **259**, 4629-4633 (1984).
82. Nambi, P., Peters, J. R., Sibley, D. R. & Lefkowitz, R. J. *J. biol. Chem.* **260**, 2165-2171 (1985).
83. Benovic, J. L. et al. *J. biol. Chem.* **260**, 7094-7101 (1985).
84. Sibley, D. R., Nambi, P., Peters, J. R. & Lefkowitz, R. J. *Biochem. biophys. Res. Commun.* **121**, 973-979 (1984).
85. Kelleher, D. J., Pessin, J. E., Ruoho, A. E. & Johnson, G. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4316-4320 (1984).
86. Stadel, J. M., De Lean, A., Mullikin-Kilpatrick, D., Sawyer, D. D. & Lefkowitz, R. J. *J. Cyclic Nucleotide Res.* **7**, 37-47 (1981).
87. Noda, C. et al. *J. biol. Chem.* **259**, 7747-7754 (1984).
88. Koschel, K. *Eur. J. Biochem.* **108**, 163-169 (1980).
89. Attramadal, H., Le Gac, F., Jahnsen, T. & Hansson, V. *Molec. cell. Endocr.* **34**, 1-6 (1984).
90. Leeb-Lundberg, L. M. F. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
91. Liebman, P. A. & Sitaramayya, A. *Adv. Cyclic Nucleotide Prot. Phosphor. Res.* **17**, 215-225 (1984).
92. Shichi, H. & Somers, R. L. *J. biol. Chem.* **253**, 7040-7046 (1978).
93. Somers, R. L. & Klein, D. C. *Science* **226**, 182-184 (1984).
94. Shichi, H., Yamamoto, K. & Somers, R. L. *Vision Res.* (in the press).
95. Engstrom, P. & Hazebauer, G. L. *Cell* **20**, 165-171 (1980).
96. De Franco, A. L. & Koshland, D. E. Jr *Proc. natn. Acad. Sci. U.S.A.* **77**, 2429-2434 (1980).
97. Goy, M. F., Springer, M. S. & Adler, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4964-4969 (1977).

Milankovitch forcing of fluctuations in the level of tropical lakes from 18 to 0 kyr BP

John E Kutzbach

Center for Climatic Research, University of Wisconsin-Madison, 1225 West Dayton Street, Madison, Wisconsin 53706, USA

F. Alayne Street-Perrott

Tropical Palaeoenvironments Research Group, School of Geography, Mansfield Rd, Oxford OX1 3TB, UK

An atmospheric general-circulation model has been used to simulate the climates of January and July at 3,000-yr intervals. From 15 kyr BP onwards, the model simulates a strengthened monsoon circulation and increased precipitation in the Northern Hemisphere tropics, culminating at 9–6 kyr BP. The computed hydrological budgets were used to estimate the area of closed lakes between 8.9 and 26.6° N, giving, with minor exceptions, results in good agreement with geological evidence.

FLUCTUATIONS in the depth and area of closed lakes (lakes without surface outlets) provide the most widespread, best dated and most spectacular evidence for late Quaternary climatic changes in low latitudes^{1–7}. At the last glacial maximum, ~18 kyr BP, lake levels in the northern tropics were low or falling⁸. This drying trend continued until 12.5 kyr BP, when water levels began to rise again. At the culmination of the ensuing lacustrine phase, ~9–8 kyr BP, the Saharan, Arabian and Thar Deserts were dotted with lakes and swamps. Water levels continued to be high until ~5 kyr BP, when desiccation set in. The extent of present-day lakes is broadly comparable with 18 kyr BP.

Changes in the orbital parameters of the Earth may provide a partial explanation for the observed trend in tropical lake levels. The seasonal cycle of solar radiation was considerably amplified in the Northern Hemisphere between ~15 and 6 kyr BP because perihelion occurred in northern summer (it now falls in northern winter) and because the axial tilt was almost 1° larger than today⁹. At the time of maximum amplification, 11–10 kyr BP, insolation was ~30 W m⁻² (7%) greater than today in July and ~25 W m⁻² (7%) less in January. These changes in seasonal solar radiation were larger than the forcing perturbations typically used in sensitivity experiments: for example, a 1–2% change in the solar constant (global average 3.4–6.8 W m⁻²)¹⁰ or a doubling of atmospheric CO₂ (corresponding to a 4 W m⁻² change in the radiative heating at the Earth's surface)¹¹. Note, however, that ours is a seasonal rather than an annual mean perturbation.

Because of their different thermal properties, Northern Hemisphere continents and ocean surfaces should have responded very differently to the increased amplitude of the seasonal radiation cycle. The large heat-storage capacity (thermal inertia) of the oceanic mixed layer should have damped the seasonal response of sea-surface temperatures (SST) to the amplified solar-radiation cycle, whereas the land-surface temperature response would have been large. For a 100-m mixed-layer ocean, the expected temperature change is negligible for purposes of sensitivity experiments (~±0.5 K); for land surfaces, it is much larger, about ±5 K (warmer in July and colder in January in both cases). From hydrostatic considerations, these temperature changes should have caused lower atmospheric pressures over land and higher pressures over the oceans in July, with the opposite changes in January. In other words, the amplified seasonal cycle of solar radiation in the Northern Hemisphere should have amplified the normal monsoon circulations, giving rise to stronger summer and winter monsoons. Experiments with atmospheric general-circulation models (AGCMs) for the radiation conditions at 9 kyr BP have confirmed this enhanced monsoonal response to changed orbital forcing and indicated substantial increases in precipitation (*P*) and precipitation minus evaporation (*P* – *E*) in the Northern Hemisphere tropics and subtropics around 9 kyr BP (refs 12–14). In contrast, the

Southern Hemisphere exhibited reduced seasonality of solar forcing (compared with the present), because perihelion was then in southern winter.

Experimental design

We have conducted a more extensive series of palaeoclimatic experiments with the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM) for 18, 15, 12, 9, 6 and 3 kyr BP (ref. 15), to study the evolution of climate since the last glacial maximum. (A detailed description and evaluation of the CCM and its simulation of modern perpetual January and July climates is given in ref. 16.) The model incorporates atmospheric dynamics based on the equations of fluid motion and includes radiative and convective processes, cloudiness, precipitation and evaporation. Orographic influences of mountains are included. SST, sea-ice limits, snow cover, land albedo and wetness, and atmospheric composition are prescribed. The model has nine levels in the vertical and an effective horizontal resolution of 4.4° lat. by 7.5° long.

The experimental design for the palaeoclimatic simulations

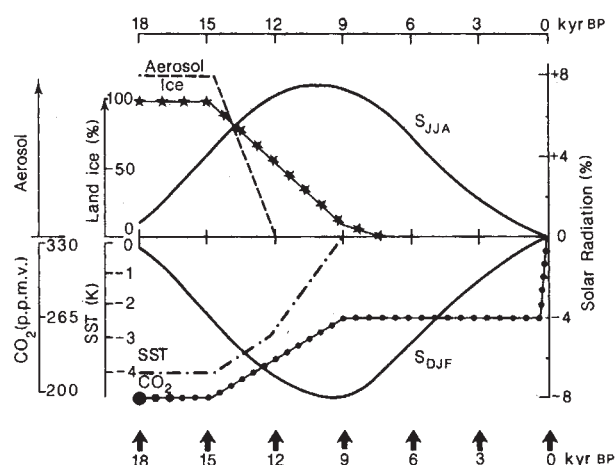


Fig. 1 Schematic diagram of major changes since 18 kyr BP in external forcing (Northern Hemisphere solar radiation in June–August (S_{JJA}) and December–February (S_{DJF}), as per cent difference from present) and internal boundary conditions (land ice¹⁷ as per cent of 18 kyr BP ice volume; global mean-annual SST¹⁷ (departure from present, K); excess glacial aerosol^{53,54}, including sea ice, on an arbitrary scale; and atmospheric CO₂ concentration^{18,19} (p.p.m.v.)). Arrows correspond to the seven sets of simulation experiments with the CCM. One experiment for 18 kyr BP included the lowered CO₂ concentration (200 p.p.m.v., large solid circle); the main series of experiments used the same CO₂ concentration as the control case (330 p.p.m.v.) rather than the actual stepwise increase¹⁹.

Table 1 Area averaged values of ($P - E$) in control cases and in palaeoclimatic experiments

	CCM version with fixed soil moisture					CCM version with interactive soil moisture	
	No ice sheet		With ice sheet			No ice sheet	With ice sheet
Length of integration (days)	1 × 90	3 × 90	1 × 90	3 × 90	3 × 90	3 full annual cycles	1 full annual cycle
CO ₂ (p.p.m.v.)	330	330	330	330	200	330	330
0 kyr BP (control)							
Summer (mm day ⁻¹)†		1.3				1.4	
Winter (mm day ⁻¹)‡		-1.2				-0.1	
Annual average (mm day ⁻¹)		0.1				0.4	
Annual total (mm)		28				149	
9 kyr BP							
Summer (mm day ⁻¹)	2.7**		2.6*	2.0**			2.3**
Winter (mm day ⁻¹)				-0.8***			-0.1
Annual average (mm day ⁻¹)			(0.9)*	0.6***			0.7*
Annual total (mm)			(331)*	227***			270*
18 kyr BP							
Summer (mm day ⁻¹)			1.3	1.3	1.1		
Winter (mm day ⁻¹)				-0.9**	-0.9**		
Annual average (mm day ⁻¹)			(0.2)	0.2	0.1		
Annual total (mm)			(68)	67	38		

The estimated annual ($P - E$) based on July 1 × 90 and January 3 × 90-day integrations is shown in parentheses. Significance of differences from control case (two-sided t -test): * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† Fixed soil-moisture version, July. Interactive soil-moisture version, June/July/August.

‡ Fixed soil-moisture version, January. Interactive soil-moisture version, December/January/February.

is shown in Fig. 1. Appropriate changes were made in both external and internal boundary conditions. The solar radiation in the model was set to past values by adjusting the eccentricity, tilt and time of perihelion. The solar-radiation values shown in Fig. 1 are average departures for summer and winter over the Northern Hemisphere. Changes in the Southern Hemisphere were of opposite sign.

The surface-boundary conditions were varied between experiments as follows. For 18 and 15 kyr BP, ice-sheet locations and heights, sea-ice limits, land-surface albedos, SST and sea level follow ref. 17. In one experiment, the atmospheric CO₂ concentration at 18 kyr BP was also lowered from its present concentration of 330 p.p.m.v. (parts per 10⁶ by volume) to its estimated glacial level of 200 p.p.m.v. (refs 18, 19). From 15 kyr BP onwards, the CLIMAP boundary conditions were altered to reflect the known patterns of deglaciation. These changes are indicated on Fig. 2. At 12 kyr BP, the extent of the North American and Eurasian ice sheets was reduced and their volume set at ~40% of glacial maximum. The sea level was raised by the appropriate amount of meltwater addition. The sea-ice edge was positioned half-way between 18–15 kyr BP and modern limits. SST anomalies were reduced to half of their 18–15 kyr BP values, except in the North Atlantic, where they were kept unchanged to approximate the cold conditions that resulted from the addition of glacial meltwaters²⁰. Land albedo was set to modern values at 12 kyr BP, except over the residual North American and European ice sheets. In the 9 kyr BP experiment, the only surface-boundary condition that differed from today was the small ice sheet over eastern North America (~10% of the full glacial volume). For 6 and 3 kyr BP, the surface-boundary conditions were identical to the control simulation (0 kyr BP).

Our January and July control experiments and all of the January palaeoclimatic experiments consisted of 450 days of simulation. A typical 450-day record was divided into three 90-day segments separated by 60-day segments. The three independent 90-day averages were then combined to form an 'ensemble mean'. For each variable, the difference between the ensemble mean of each palaeoclimatic experiment and the control run was tested for statistical significance using a measure of the natural variability of the model derived from the 6 individual 90-day averages²¹. Our main series of July palaeoclimatic experiments consisted of 150-day integrations, from which

one 90-day segment was extracted and averaged; in these cases, the natural variability of the model was estimated only from the control experiment. The shorter runs were possible because the natural variability is lower in July than in January in the Northern Hemisphere. However, to test whether or not the results of these shorter July experiments would differ significantly from the results of longer simulations, 450-day July experiments were run for 9 and 18 kyr BP and compared with the 150-day experiments (Table 1). The longer integrations produced similar results but the statistical significance of the results was generally higher.

The perpetual January and July averages, appropriately weighted for the length of the astronomical seasons at each time period (18–0 kyr BP), were averaged to approximate the annual mean. In more lengthy simulation experiments, in which the entire seasonal cycle was simulated, we found the true annual mean and the annual mean estimated from the average of January and July simulations to be in close agreement. (For a detailed summary of the model characteristics, computational and statistical procedures, boundary conditions and simulated global palaeoclimatic patterns, see refs 14, 15.)

Although the CCM has a spatial resolution sufficient, in principle, to define the regional details of climatic change, we have chosen to present the area-averaged results for an array of 55 land grid points in the Northern Hemisphere tropics (8.9–26.6° N) (Fig. 2). This latitudinal belt encompasses the region of most coherent lake-level anomalies^{5,6,8}. From the point of view of the natural variability of the model, averaging over this large area leads to much more stable and statistically significant results than an examination of individual grid squares. Of the 55 grid points, 37 (67%) are located in Africa, 15 in South and South-east Asia, and 3 in Central America (Fig. 2).

A terminal lake responds to an increase in runoff by expanding in area. Provided that interbasin transfers of groundwater are negligible, equilibrium is established when

$$z = A_L/A_B = R/(E_L - P_L) \quad (1)$$

where z is the lake-area ratio, A is area, R is runoff, E is evaporation, P is precipitation and the subscripts L and B refer to the lake and its drainage basin, respectively^{22–24}. In arid and semiarid areas, $E_L \gg E_B$. At equilibrium, $R = P_B - E_B$.

The equilibrium extent of closed lakes in the study area at each time period was calculated from the model-generated area

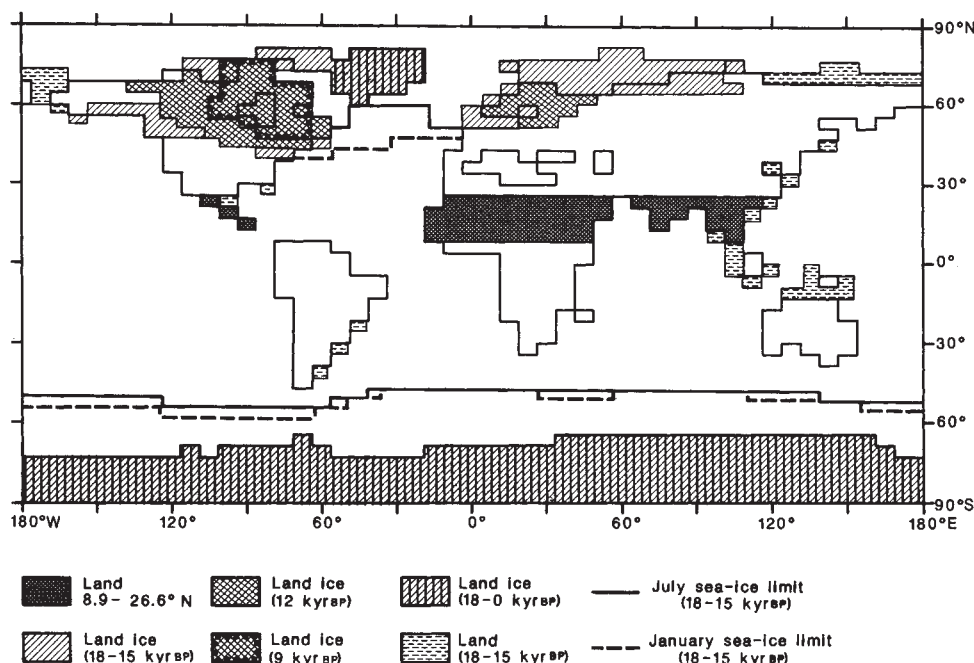


Fig. 2 Selected surface-boundary conditions for the palaeoclimatic simulation experiments: limits of land ice at 18, 15, 12, 9 and 0 kyr BP; July and January sea-ice limits at 18 and 15 kyr BP; and areas of exposed continental shelf at 18 and 15 kyr BP. (For simplicity, we have not shown the prescribed changes in SST, land albedo and ice-sheet topography, and the sea-ice limits for 12–0 kyr BP. See text for details.) The location of the 55 tropical land grid points (8.9–26.6° N) is also indicated.

averages of P and $(P - E)$, using the modified equation

$$z = (1/e)(P - E)/(E_L - P) \quad (2)$$

where e is the ratio of total land area to the area of internal drainage (endorheic area) within this belt. We assume that $P \approx P_L \approx P_B$. As a first approximation, in view of the paucity of runoff data from the subtropical deserts, the adjustable parameter e was introduced to partition the runoff between the areas of internal drainage in which closed lakes can form and the areas draining to the ocean. The modern area-average value of e for the study area is 2.6 (ref. 25). Thus, for the modern simulation, $\sim 38\%$ ($1/2.6$) of the model-generated area-average $(P - E)$ is available for runoff to closed interior basins. No attempt has been made to estimate the z ratio for lakes within the areas of external drainage, to which equation (1) is not applicable. In fact, many additional lakes developed during the early and mid-Holocene along river systems draining into the ocean^{26,27}.

The CCM does not calculate E_L explicitly for hypothesized interior lakes. Moreover, except for large enclosed seas, such as the Mediterranean and Caspian, its coarse spatial resolution precludes explicit simulation of inland water bodies. A modern value of $E_L = 2,600 \text{ mm yr}^{-1}$ was assumed. This is representative of Lake Chad (13° N) in dry years²⁸ and is comparable with the evaporation rate calculated by the CCM for tropical ocean areas with offshore winds. It is also comparable to the rate obtained if the model-generated winds and mixing ratios over land (8.9–26.6° N) are adjusted for flow over a water surface with a drag coefficient intermediate between land and open ocean. In the palaeolake simulations, E_L has been varied in proportion to the calculated changes in land evaporation, to allow for variations in surface temperature, humidity and wind speed.

Results

The area-averaged CCM results and their statistical significance are summarized in Figs 3–4, and Table 1. In July, the net radiation increases from 164 W m^{-2} at 18 kyr BP to a peak of 181 W m^{-2} in the 9 kyr BP experiment (12% greater than the control) and then decreases again (Fig. 3a). In January it varies in the opposite direction, reaching its minimum (-16%) at 9 kyr BP.

The surface-temperature lowering at 18 kyr BP is 3.5 K in January and 2.0 K in July (Fig. 3b). Although the estimated mean-annual temperature rises rapidly after 12 kyr BP and does not differ significantly from the control case in the 9, 6 and 3 kyr BP experiments, the seasonal temperature range is 2–4 K greater

than today between 15 and 6 kyr BP, and 3–4 K greater at 12 and 9 kyr BP. The highest July temperature departure ($+1.3 \text{ K}$) occurs in the 9 kyr BP experiment.

The simulated changes in the January surface wind field are small. In July, the CCM indicates weak northwesterly flow at 18 and 0 kyr BP (Fig. 3c). In marked contrast, the wind direction at 15, 12, 9 and 6 kyr BP is southwesterly. This intensified low-level monsoon flow, consistent with lower pressure over the continents and higher pressure over the oceans, reaches its greatest strength and statistical significance at 9 kyr BP, implying an increased transport of moisture from the oceans onto the Northern Hemisphere land masses.

Figure 4 summarizes the simulated hydrological cycle for each experiment. In July, the precipitation rate is lowest at 18 and 0 kyr BP, attaining its maximum values in the 9 and 6 kyr BP simulations (2.1 and 18 mm day^{-1} greater than the control value of 5.6 mm day^{-1} , respectively, representing 30–35% increase in precipitation ($p < 0.05$). Although the land evaporation rate is also greater ($p < 0.01$) at these times, as a result of higher summer

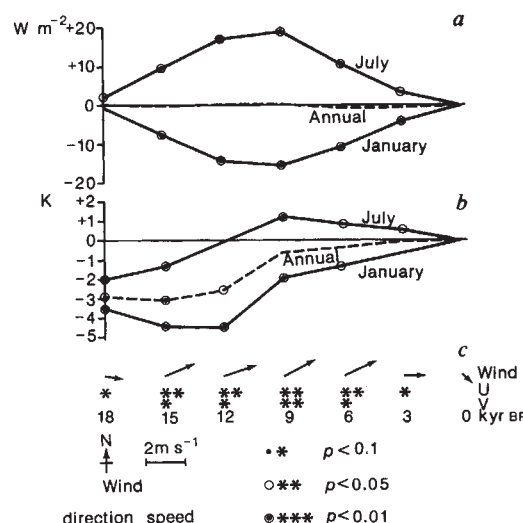


Fig. 3 Simulated climate of the land in the latitude belt 8.9–26.6° N, 18–0 kyr BP. a, Net radiation, expressed as departures from the modern control case; b, surface temperature, ditto; c, July wind, arrows. Statistical significance of the results (two-sided t -test)²¹ is shown by symbols. In c, significance of the U (zonal) and V (meridional) wind components is given separately.

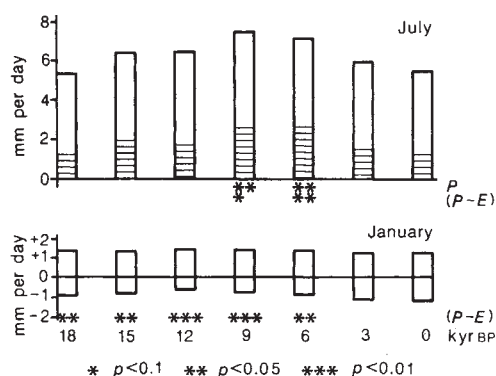


Fig. 4 Simulated hydrological budgets for July and January, land surface, in the latitude belt 8.9–26.6° N, 18–0 kyr BP. For each experiment, top of column represents P , unshaded area E , and shaded area, if any, $(P-E)$ available for runoff. Statistical significance of P and $(P-E)$ (two-sided t -test)²¹ is shown by symbols.

temperatures and wind speeds, the variations in evaporation are much smaller, so that $(P-E)$ is largest at 6 and 9 kyr BP (>2.6 mm day⁻¹).

In January, the dry season in the Northern Hemisphere tropics, the changes in precipitation are not statistically significant. However, land evaporation is significantly less in the palaeoclimatic experiments than in the control case ($p < 0.05$ or $p < 0.01$), due to lower winter temperatures (and weaker zonal wind speeds). As a result, the net precipitation deficit is significantly reduced in all cases except 3 kyr BP, with maximum reductions at 12 and 9 kyr BP (0.6 and 0.4 mm day⁻¹ respectively), compared with the control.

Figure 5a shows the simulated values of annual P and $(P-E)$. At 9 and 6 kyr BP, P increases by 300–350 mm yr⁻¹ (25–30%) compared with the control. It shows a general increasing trend from 18 to 9–6 kyr BP, followed by a decrease to the present (the modern control value is higher than observed)^{16,25}. However, the calculated value of $(P-E)$ for the control case (28 mm yr⁻¹) falls midway between the estimates derived from modern hydrological²⁵ and aerological²⁹ data. Mainly because of the highly favourable combination of increased rainy-season precipitation and reduced evaporation during the dry season, $(P-E)$ is much larger (~ 330 mm yr⁻¹) in the 9 and 6 kyr BP experiments, so that both P and $(P-E)$ results are in good accord with independent palaeoclimatic estimates ($\Delta P = 130$ –505 mm yr⁻¹; refs 30, 31). In contrast, the values for 18 kyr BP and 3 kyr BP appear statistically indistinguishable from the present.

The reliability of these conclusions is reinforced by the results of the additional CCM experiments summarized in Table 1. These involved tests of the sensitivity of the July results for 9 and 18 kyr BP to the length of integration (see above), and of the sensitivity of the July climate to the presence of the residual North American ice sheet at 9 kyr BP. As far as the hydrology of the tropics is concerned, the experiment for 9 kyr BP without the ice sheet gave almost identical results to the one with the ice sheet. We conclude that the results are not strongly dependent on the exact choice of high-latitude boundary conditions. An experiment for 9 kyr BP was also made using a version of the CCM with interactive soil moisture and a full seasonal cycle. This model produced a comparable increase in Northern Hemisphere summer-monsoon precipitation, although the net changes in mean-annual P and $(P-E)$ were smaller, because of the wetter winter conditions in the control run.

The results of the various CCM experiments for 9 kyr BP (Table 1) are also generally consistent with results obtained with two other AGCMs: a University of Wisconsin spectral model of relatively low horizontal resolution (about 12° lat. by 11° long.) and 5 vertical levels, integrated for a full annual cycle with 9-kyr BP solar radiation^{12,13}; and a British Meteorological Office grid-point model of horizontal resolution comparable

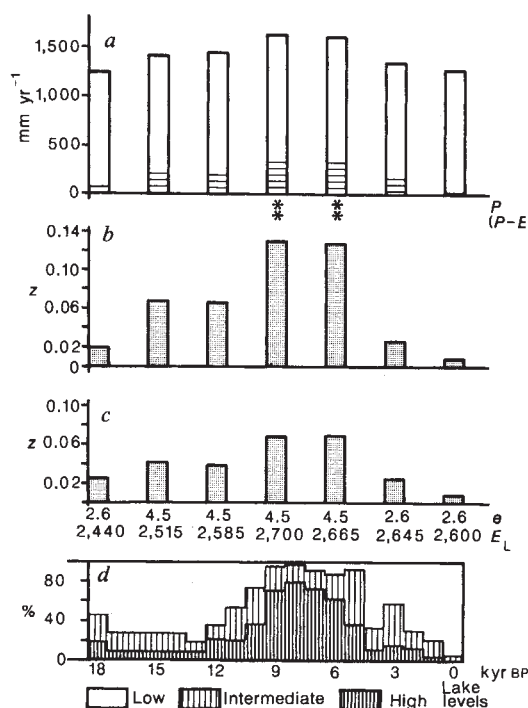


Fig. 5 Comparison of simulated hydrological budgets, simulated z ratios and lake-level evidence for the latitude belt 8.9–26.6° N, 18–0 kyr BP. *a*, P and $(P-E)$ (key as for Fig. 4); *b*, z ratios assuming constant e , E_L ; *c*, z ratios assuming variable e , E_L ; *d*, temporal variations in percentage of lakes with low, intermediate or high⁶ levels. (Maximum sample size = 73.) From Oxford lake-level data bank.

with the CCM (3° lat. by 5° long.) and 5 vertical levels, integrated for 30 days with June 10 kyr BP solar radiation^{32,33}. However, the discussion of the latter experiment emphasized climatic changes at high latitudes rather than in the tropics or subtropics.

Figure 5b shows the z ratio obtained by applying equation (2) to the CCM results for each time period, assuming present-day values of e and E_L . The calculated modern value of z is small (<0.01), in reasonable agreement with ref. 34. This simple lake-area model yields a low z ratio at 18 kyr BP and highest ratios at 9 and 6 kyr BP. However, the magnitude of z at these times seems to be unrealistically large.

In Fig. 5c, e and E_L have been allowed to vary between experiments in a way consistent with the model climate. At present, e decreases with increasing aridity from 5.6 at 5–10° N to 2.0 at 20–25° N (ref. 25). To approximate the feedback effect of climatically-induced variations in the drainage network, e was adjusted in line with the simulated latitudinal shifts in the July Inter Tropical Convergence Zone (ITCZ). E_L was varied in proportion to the computed changes in land evaporation. The values of e and E_L used are shown on Fig. 5c. The sensitivity of the results to these assumptions is given by $\Delta z = -z\Delta e/e$ and $\Delta z = -z\Delta E_L/E_L$.

In Fig. 5c, the extent of lakes at 18 kyr BP remains small. The magnitude of z at 9 and 6 kyr BP is reduced to more realistic values, caused by the combined effects of increased evaporation and smaller endorheic area. The results in Fig. 5c are in excellent qualitative agreement with the sequence of lake-level fluctuations in the latitude belt 8.9–26.6° N (Fig. 5d), except for 15 kyr BP and, to a lesser extent, 12 kyr BP. While the observations suggest that the water-level rise was underway by 12 kyr BP, 15 kyr BP clearly falls within the phase of lowest levels. Unfortunately, there is a lack of spatially representative lake-area data for past periods, so it is not yet possible to check quantitatively the simulated z ratios.

As a result of the substantial increase in northern summer radiation beginning with the 15-kyr BP experiment (Figs 1, 3), land temperatures rise and increased inflow (Fig. 3c) and vertical

motion produce increased precipitation. Hence, the CCM generates much more runoff at 15 and 12 than at 18 kyr BP. The net effect is a large area of lakes (Fig. 5b, c). Although the increase in annual ($P-E$) does not reach a high level of statistical significance ($p < 0.2$), the disagreement between the lake-area model and the observed lake-level record for 15 and 12 kyr BP is interesting. Such a discrepancy (equivalent to ~ 1 – 1.5 mm day $^{-1}$ summer precipitation), if real, could result from inaccuracies in the reconstructed lake-level record, deficiencies in either the AGCM or the z -ratio model, and/or other types of forcing not included in the experimental design.

There is abundant independent evidence for aridity, reduced plant cover and equatorward spread of sand dunes in the study area during the interval 18–12.5 kyr BP. Although runoff from the Saharan mountains began to increase around 16 kyr BP (refs 3, 35–42), most lowland basins continued to experience low water levels during the next 3.5 kyr (ref. 8); a lag which seems too long to be explicable solely in terms of non-climate factors such as delayed aquifer response.

Two internal features of the CCM may cause precipitation over tropical arid and semiarid land areas to be overestimated. The version used in our main series of experiments calculates evaporation using a bulk formula⁴⁴ with fixed values of ground wetness, C_w , and surface drag, C_D , for all land surfaces. Black and Pitcher⁴⁵ found that the precipitation over northern Africa decreased by 1–2 mm day $^{-1}$ if C_w was reduced to values typical of desert and semidesert. The Goddard Laboratory for Atmospheres AGCM exhibits similar sensitivity to a decrease in surface roughness, its response taking the form of a reduction in cross-isobaric moisture convergence and precipitation over tropical land areas^{46,47}. Hence, an even better fit to the geological record might be obtained by allowing C_w and C_D to vary more realistically in the CCM experiments.

Additional experiments

Two major types of forcing not included in our main series of experiments are variations in atmospheric CO_2 and aerosol content. Table 1 shows the results of the experiment for 18 kyr BP in which atmospheric CO_2 was reduced to 200 p.p.m.v. The outcome was a small decrease in annual-average ($P-E$) (30 mm yr $^{-1}$) over the study area. In the CCM, the changed CO_2 concentration led to lower land temperatures, SST being prescribed following CLIMAP. By lowering temperatures over land relative to the oceans, the summer-monsoon circulation was

weakened. (If Oeschger *et al.*⁴⁸ are correct, and the recovery from low glacial CO_2 levels did not begin until around 13 kyr BP, then this mechanism would help to account for the observed aridity. In addition to the direct climatic effects modelled here, however, the stomatal response of plants to lower CO_2 (ref. 49) also needs to be considered. Another important task is to assess the climatic impact of the high tropospheric loadings of terrigenous and marine aerosols characteristic of glacial times^{50–54}, which may have led to increased atmospheric stability and reduced convective precipitation.

Conclusions

The general agreement between our results and the geological record suggests that the Milankovitch mechanism has been largely responsible for long-term variations in lake levels in the Northern Hemisphere tropics and subtropics since 18 kyr BP, and that the influence of changes in boundary conditions associated with glaciation in high latitudes has been comparatively minor. The simulated changes in winds, P and ($P-E$), can be broadly characterized as an enhanced seasonal (monsoonal) response of the Northern Hemisphere climate to the amplified seasonal cycle of solar radiation in that hemisphere. This response is largest and of greatest statistical significance at 9 and 6 kyr BP. Even the magnitude of the model response at these times agrees with existing geological evidence. However, note that variations in the Earth's orbit have occurred on too long a timescale to account for the shorter hydrological fluctuations (lasting 10^2 – 10^3 yr) superimposed on the broad trends shown in Fig. 5d (refs 6, 7, 55–57). A convincing explanation for these shorter-term events remains to be found.

We thank Peter Guetter for running the experiments and processing the results, and Neil Roberts and Sandy Harrison for compiling the lake-level data. This research was carried out while F.A.S.-P. was a visiting associate professor at the Center for Climatic Research, Institute of Environmental Studies, and Department of Geography, University of Wisconsin-Madison, and was supported by the NSF Climate Dynamics Program (ATM-8219079 and -8412958). The model computations were made at NCAR, which is sponsored by NSF, with a computing grant from the NCAR Computing Facility (35381017). Funding for data compilation was provided by the US Department of Energy Carbon Dioxide Research Division (DE-ACO2-79EV10079). We also thank F. Bryan and A. Oort for supplying unpublished aerological data.

Received 24 April; accepted 24 June 1985.

1. Faure, H. *Mitt. int. Verein. theor. angew. Limnol.* **17**, 131–146 (1969).
2. Butzer, K. W. *et al. Science* **175**, 1069–1076 (1972).
3. McClure, H. A. in *Quaternary Period in Saudi Arabia* Vol. 1 (eds Al-Sayari, S. S. & Zötl, J. G.) 252–253 (Springer, Vienna, 1978).
4. Limbrey, S. in *Geochronology* (eds Davidson, D. A. & Shackley, M. L.) 213–226 (Duckworth, London, 1976).
5. Street, F. A. & Grove, A. T. *Quat. Res.* **12**, 83–118 (1979).
6. Street-Perrott, F. A. & Roberts, N. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. A. & Ratcliffe, R. A. S.) 331–345 (1983).
7. Wasson, R. J., Smith, G. I. & Agrawal, D. P. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **46**, 345–372 (1984).
8. Street-Perrott, F. A. & Harrison, S. P. *Geophys. Monogr.* **29**, 118–129 (1984).
9. Berger, A. L. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
10. Wetherald, R. T. & Manabe, S. *J. Atmos. Sci.* **32**, 2044–2059 (1975).
11. Ramanathan, V. *J. Atmos. Sci.* **38**, 918–930 (1981).
12. Kutzbach, J. E. *Science* **214**, 59–61 (1981).
13. Kutzbach, J. E. & Otto-Bliesner, B. L. *J. Atmos. Sci.* **39**, 1177–1188 (1982).
14. Kutzbach, J. E. & Guetter, P. J. in *Milankovitch and Climate*, part 2 (eds Berger, A. L. *et al.*) 801–820 (Reidel, Dordrecht, 1984).
15. Kutzbach, J. E. & Guetter, P. J. *Ann. Glaciol.* **5**, 85–87 (1984).
16. Pitcher, E. J. *et al. J. Atmos. Sci.* **40**, 580–604 (1983).
17. CLIMAP Project Members *Geol. Soc. Am. Map Chart Ser. MC-36* (1981).
18. Neftel, A. *et al. Nature* **295**, 220–223 (1982).
19. Lorius, C. *et al. Ann. Glaciol.* **5**, 88–94 (1984).
20. Ruddiman, W. F. & McIntyre, A. *Science* **212**, 617–627 (1981).
21. Chervin, R. M. & Schneider, S. H. *J. Atmos. Sci.* **33**, 413–423 (1976).
22. Snyder, C. T. & Langbein, W. B. *J. geophys. Res.* **67**, 2385–2394 (1962).
23. Bowler, J. M. *Hydrobiologia* **82**, 431–44 (1981).
24. Street-Perrott, F. A. & Harrison, S. P. in *Paleoclimate Data and Modeling* (ed. Hecht, A. D.) 291–240 (Wiley, New York, 1985).
25. Baumgartner, A. & Reichel, E. *The World Water Balance—Mean Annual Global, Continental and Maritime Precipitation* (Elsevier, Amsterdam, 1975).
26. Grove, A. T. & Warren, A. *Geogr. J.* **134**, 194–208 (1968).
27. Williams, M. A. J. & Adamson, D. A. (eds) *A Land Between Two Niles: Quaternary Geology and Biology of the Central Sudan* (Balkema, Rotterdam, 1982).
28. Tetzlaff, G. & Adams, L. J. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. A. & Ratcliffe, R. A. S.) 347–360 (Reidel, Dordrecht, 1983).
29. Bryan, F. & Oort, A. H. *J. geophys. Res.* (in the press).
30. Street, F. A. *Palaeoecol. Afr.* **11**, 135–143 (1979).
31. Kutzbach, J. E. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. A. & Ratcliffe, R. A. S.) 371–389 (Reidel, Dordrecht, 1983).
32. Mason, B. J. *J. R. Met. Soc.* **102**, 473–498 (1976).
33. Mitchell, J. F. B. *Met. O 20 Tech. Note II/100* (Meteorological Office, Bracknell, 1977).
34. Schuling, R. D. in *Interactions Between Sediments and Freshwater* (ed. Golterman, H. L.) 12–18 (PUDOC, Wageningen, 1977).
35. Grove, A. T. *Geogr. J.* **124**, 526–533 (1958).
36. Pastouret, L. *et al. Oceanol. Acta* **1**, 217–232 (1978).
37. Rossignol-Strick, M. & Duzer, D. *Pollen Spores* **21**, 105–134 (1979).
38. Talbot, M. R. in *The Sahara and the Nile* (eds Williams, M. A. J. & Faure, H.) 37–62 (Balkema, Rotterdam, 1980).
39. Watts, W. A. & Bradbury, J. P. *Quat. Res.* **17**, 56–60 (1982).
40. Wasson, R. J. *et al. Zeits. Gemorph. Suppl.* **45**, 117–151 (1983).
41. Leyden, B. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4856–4859 (1984).
42. Talbot, M. R. *Palaeoecol. Afr.* **16**, 203–214 (1984).
43. Jäkel, D. *Palaeoecol. Afr.* **11**, 13–39 (1979).
44. McAvaney, B. J., Bourke, W. & Puri, K. *J. Atmos. Sci.* **35**, 1557–1583 (1978).
45. Black, M. L. & Pitcher, E. J. *Abstr. 3rd Conf. on Climate Variations and Symposium on Contemporary Climate* 91–92 (American Meteorological Society, 1985).
46. Sud, Y. C., Shukla, J. & Mintz, Y. *Abstr., 3rd Conf. on Climate Variations and Symposium on Contemporary Climate* 93–94 (American Meteorological Society, 1985).
47. Sud, Y. C. & Smith, W. E. *Abstr., 3rd Conf. on Climate Variations and Symposium on Contemporary Climate* 128–129 (American Meteorological Society, 1985).
48. Oeschger, H. *et al. in Palaeoclimatic Models and Research* (ed. Ghazi, A.) 95–107 (Reidel, Dordrecht, 1983).
49. Wigley, T. M. L. & Jones, P. D. *Nature* **314**, 149–152 (1985).
50. Bowler, J. M. in *Antarctic Glacial History and World Palaeoenvironments* (ed. van Zinderen Baker, E. M.) 149–172 (Balkema, Rotterdam, 1978).
51. Kolla, V. & Biscaye, P. E. *J. sedim. Petrol.* **47**, 642–649 (1977).
52. Kolla, V., Biscaye, P. E. & Hanley, A. F. *Quat. Res.* **11**, 261–277 (1979).
53. Petit, J. R., Brist, M. & Royer, A. *Nature* **293**, 391–394 (1981).
54. Thompson, L. G. & Mosley-Thompson, E. *J. volcan. geotherm. Res.* **11**, 11–27 (1981).
55. Bradbury, J. P. *et al. Science* **214**, 1299–1305 (1981).
56. Rognon, P. in *Palaeoclimatic Research and Modelling* (ed. Ghazi, A.) 114–123 (Reidel, Dordrecht, 1983).
57. Gillespie, R., Street-Perrott, F. A. & Switsur, R. *Nature* **306**, 680–683 (1983).

Uplift and extension at the Gulf of Suez: indications of induced mantle convection

Michael S. Steckler

Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

Reconstruction of tectonic movements since the initiation of the Gulf of Suez rift indicates that there has been 25–27 km of extension. The uplift bordering rift indicates that lithosphere heating greatly exceeds that producible by uniform lithosphere extension. I propose that small-scale convection induced by the rifting augments the heating introduced by extensional rifting and produces the broad uplifts flanking the rift.

CONTINENTAL margins and rift basins are currently believed to have been formed by lithosphere extension. The vertical motions that result from uniform extension have been elegantly quantified by McKenzie¹. The two main contributions to these motions were shown to be subsidence caused by crustal thinning and uplift caused by lithosphere heating. For a given extension factor (with assumed crustal and lithosphere thicknesses) the net subsidence or uplift is prescribed. This model has been successful at explaining the first-order observations at extensional sedimentary basins. Several authors^{2–5} have expanded on the initial model to fit observed subsidence patterns.

Lithosphere extension has generally been modelled as a passive process in which all heating occurs as the result of the extension and no asthenosphere heat sources are included. This results in an inability, even when two-dimensional effects are included, to produce uplifts on the sides of rifts >4–500 m elevation.

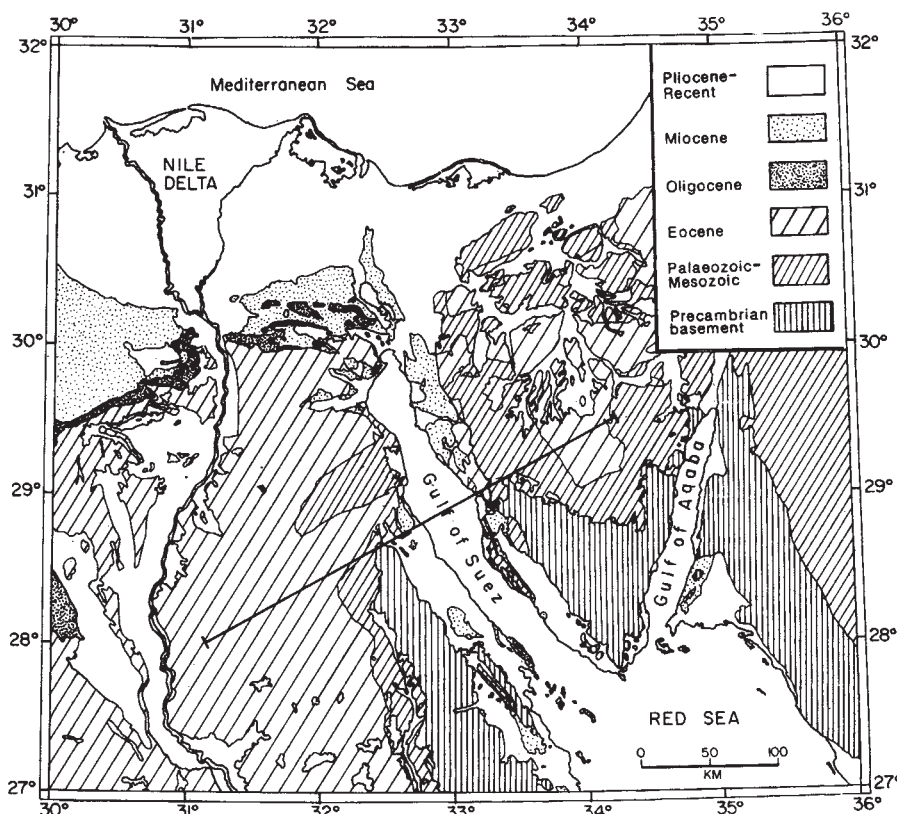
Observations of rift flank uplifts of ≥ 1 km are commonly observed^{6–10}. An active asthenosphere heat source is often assumed to produce the uplift and act as a driving mechanism for the rifting^{11–13}. In an active rift system driven by a deep heat

source, uplift should precede rifting. The question of the temporal and causal relationship between uplift and rifting has been debated for some time^{14–18} and no clear consensus exists. In the Gulf of Suez, the timing and magnitude of the uplift and subsidence produced by rifting can be documented and the contribution of lithosphere extension calculated. Thus, the Gulf of Suez provides a test of proposed rifting models.

Geology of the Gulf of Suez

The Gulf of Suez is the northwestern branch of the Red Sea (Fig. 1). The rift is 60–100 km wide and extends for 400 km, separating Africa from the Sinai Peninsula. Rifting is post-Eocene^{19,20} and was underway by the start of the Miocene^{20,21}. Initially, the Gulf of Suez probably formed as the northern part of the Red Sea rift. Subsequently, the formation of the Levant transform through the Gulf of Aqaba has accommodated 105 km of the relative motion between Arabia and Africa^{22,23}. This scenario is borne out by the alignment of the rift borders of the Gulf of Suez and Red Sea using Freund's²⁴ reconstruction and Esso's exploration²⁵ in the northern Red Sea which suggests that "a continuous sedimentary basin extended from the Gulf

Fig. 1 Geological map of the Gulf of Suez region. Heavy line indicates location of profile analysed.



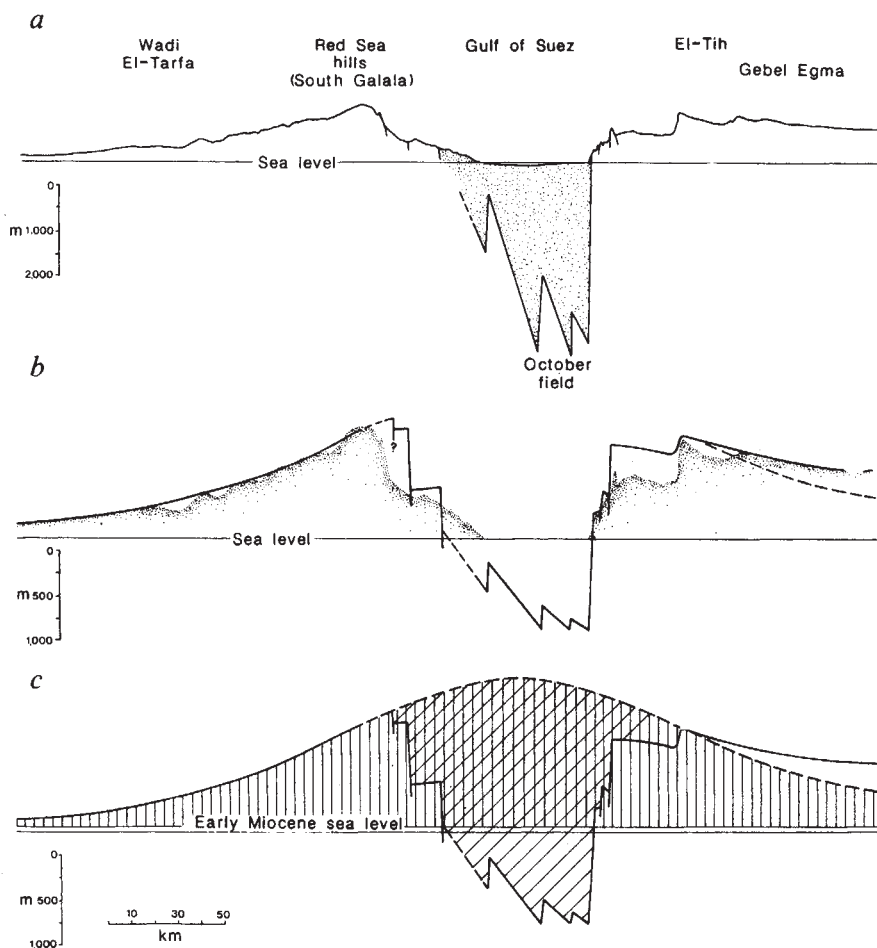


Fig. 2 *a*, Profile across the central Gulf of Suez. Shaded pattern indicates the syn-rift sedimentary fill. All the uplift and subsidence has occurred since the Eocene when shallow water carbonates were deposited in the region. Vertical exaggeration, 20:1. *b*, Reconstruction of the tectonic uplift and subsidence of the profile in the absence of erosion and sedimentation. Dashed line shows the presumed topographical fall-off in the absence of the Gulf of Aqaba to the east. Note change in vertical exaggeration to 40:1. *c*, Interpretation of the tectonic movements as uplift caused by thermal expansion of the lithosphere (vertical lines) and down-dropping of the rift basin away from this curve caused by crustal thinning (diagonal lines).

of Suez to the Red Sea". The question of whether the opening of the Gulf of Aqaba has entirely²⁶ or partially supplanted the Gulf of Suez is still unresolved, although the offset of the central depression of the far northern Red Sea towards the Gulf of Suez suggests that it has not²⁷.

The early history of rifting indicates that the large uplifts bordering the rifts were not formed as a precursory doming event, but developed during the main phase of development of the rift. During the Eocene, the entire Gulf of Suez region was a shallow water carbonate platform. The absence of Oligocene sediments, usually cited as evidence of a dome, is best explained as a mainly non-depositional hiatus caused by a major northwards regression. The Mediterranean coast of Egypt is an old north-facing continental margin and exhibits a stratigraphic pattern similar to other old margins, that is, an extensive Eocene shelf and a major regression during the Oligocene. Non-marine Oligocene is known near Cairo and in other regions north of the Gulf of Suez (Fig. 1). The only Oligocene beds known near the Gulf of Suez are a few locations within the rift, exposed onshore at Abu Zenima (29° 1'N, 33° 11'E)^{20,28} and penetrated by wells in the northern Gulf²⁹. If an uplift had existed before any rifting, the crest of the dome is the most unlikely place to contain preserved Oligocene beds.

When rifting commenced, with the development of tilted fault blocks and minor basalt activity, major erosion of the Eocene and older rocks occurred only on the high ends of the fault blocks. The opposing ends experienced deposition of marine carbonates with local conglomerates. The simultaneous sub-aerial erosion and marine deposition on the same tilted blocks are clear evidence that the rift initiated near sea level.

Throughout the Lower Miocene, the topography remained subdued with the deposition of the Lower Rudeis formation, a monotonous shaly-marly with fairly uniform thickness throughout the Gulf²⁰. Scott and Govean²¹ estimate subsidence

rates of the order of 50 m Myr⁻¹ on the rift flanks. At mid-Rudeis time, marked by a widespread unconformity, there is a dramatic change in the depositional pattern. Subsidence rates change dramatically^{21,30} and the first widespread conglomerates appear at the rift margins²⁰. The lowermost cobbles are of the immediately pre-rift, Eocene rocks and only later do cobbles of the underlying sedimentary and Precambrian rocks appear, indicating that major uplift and unroofing began at this time²⁰.

Shortly after this event, the Gulf of Suez first became isolated from the Mediterranean and evaporitic conditions were initiated. Large thicknesses of halite and anhydrite were deposited during the Miocene until the present-day connection to the Indian Ocean was established and sedimentation returned to normal marine conditions. Throughout these periods of changing environmental conditions, the Gulf of Suez rift continued to develop, with large thicknesses of sediment and local to regional nonconformities marking the phases of the developing rift.

Tectonic analysis

The present-day state of the Gulf of Suez is illustrated in Fig. 2*a*. From west to east, the increase in elevation from the Nile Valley to the border of the rift is first seen. This is followed by rapid down-dropping into the rift where the stipple indicates the thickness of syn-rift sediments. The faults within the rift tend to dip in the same direction on any cross-section of the rift, although this direction varies along the length of the Gulf of Suez^{31,32}. Reaching the eastern side of the rift, the elevation once again rises to >1,000 m above sea level. The topography does not, however, fall off symmetrically because the Gulf of Aqaba-Dead Sea rift lies east of the section.

This cross-section has been reconstructed to show how it would appear in the absence of erosion and sedimentation (Fig. 2*b*). The pre-rift sedimentary cover consists of flat-lying Phanerozoic sequences which strike perpendicular to the present

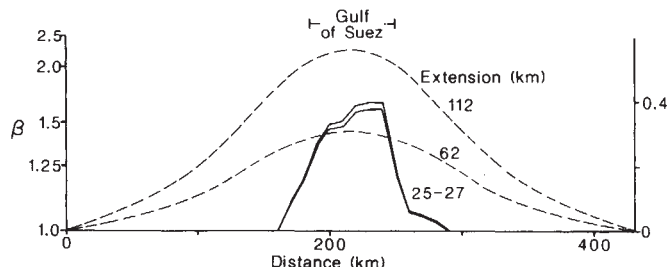


Fig. 3 Estimates of the extension factor (β) needed to produce the vertical motions estimated from Fig. 2. Scale on right-hand side gives fractional thinning ($1 - 1/\beta$) which is linear with the amplitude of vertical motion. Values to the right of the curves give total extension represented by curves. Solid lines, estimates from crustal thinning for 40- and 42-km initial crustal thickness; dashed lines, estimates from thermal uplift. Upper curve is based on 125-km-thick lithosphere and lower curve on 225-km-thick lithosphere. Heating indicates extension 2.5–4.5 times larger than crustal thinning.

Gulf of Suez. Using available data on the thickness of these units^{19,20,33}, the amount of erosion of the pre-rift cover has been calculated, the sediments restored and the elevation adjusted using Airy isostasy.

The reconstructed elevation is:

$$El_R = El_A + T_s - T_s \frac{\rho_s}{\rho_m}$$

where El_R and El_A are the restored and actual elevations, T_s is the thickness of eroded sediments and ρ_s and ρ_m the sediment and mantle densities, respectively. Using densities of 2.54 g cm^{-3} and 2.60 g cm^{-3} for the clastic Nubian sandstones and the carbonate Cenomanian–Eocene sections, the corrections to the elevations are 20–25% of the amount eroded. Thus, errors in reconstruction are minimized.

One feature of note is the Eocene outcrop on the South Galala platform and the Precambrian basement exposed to the east (either side of the westernmost fault, Fig. 2b), both restored to the same elevation. This indicates a purely erosional topographical difference across this topographical step, rather than tectonic, suggesting that the estimates of the Phanerozoic sediment thicknesses and the applicability of Airy isostasy for the restoration in this region were reasonable.

Similarly, within the rift proper, the thickness and distribution of the syn-rift sediments have been constructed from well data and structural information. The well data were used to calculate the densities, the sedimentary section was removed or backstripped and the basement adjusted upwards for the removal of the weight of these sediments³⁴. The heavy line in Fig. 2b shows the vertical tectonic movements that have occurred as a result of rifting, because prior to the initiation of the rift this region was a flat shallow sea. The dashed line shows the presumed topographical fall-off on the Sinai side assuming symmetry in the absence of the Gulf of Aqaba further east.

Because the region consisted of a shallow shelf immediately before rifting, all of the vertical movements from about sea level can be attributed rifting. The two main causes of vertical movements during rifting are heating of the lithosphere, which produces uplift, and thinning or densification of the crust, which produces a subsidence or downwards motion¹. Therefore, the profile was used to determine the extent to which these two processes have modified the lithosphere. A gaussian curve was fitted to the broad uplift of the Gulf of Suez to interpolate the uplift into the rift. Several different curves were tried; and they do not make a significant difference in the results. Fig. 2c shows the separation of the two rift components. The vertical rulings indicate the uplift resulting from thermal expansion of a heated lithosphere and the diagonal ruling shows the down-dropping away from this surface within the rift caused by the extension, faulting and thinning of the crust.

The subsidence caused by the crustal thinning is given by¹

$$S_c = C \frac{(\rho_m - \rho_c) \left[1 - (\alpha T_0/2) \frac{C}{A} \right]}{\rho_m} (1 - 1/\beta_c)$$

where ρ_c is the density of the crust, α is the coefficient of thermal expansion, T_0 the temperature at the base of the lithosphere, C and A are the crustal and lithosphere thicknesses, respectively, and β_c is the extension factor for the crust. (The values used are given in Table 1.) The term modifying the densities is caused by differential thermal expansion of the crust and mantle and is negligible. The crustal thickness of undeformed Nubian shield crust has been measured at 40–42 km in the eastern Sinai^{35,36} and Saudi Arabia³⁷. The estimates of β_c based on these crustal thicknesses and the fault-bound subsidence within the rift is given in Fig. 3. Integrating the β_c values over the rift, the total extension needed to produce the observed subsidence is 25–27 km. The maximum thinning of the crust is ~40% ($\beta = 1.65$) beneath the rift (Fig. 3).

Chenet and Letouzey³⁸ found a value of $\beta = 1.7$ for the Belayim Marine oil field (latitude $28^\circ 35'N$), which lies in a structurally equivalent position to the October field³⁹, near the extensional maximum of the profile examined here. Near the southern end of the Gulf of Suez, they estimate an overall β of 1.35, corresponding to 31.5 km of extension in the 90 km rift. As the extension decreases to the north, these values are in excellent agreement. The plate reconstructions of Freund²⁴ and Cochran^{40,41} both assume 30 km of opening in the southern Gulf of Suez. Estimating the opening on the basis of geology is difficult. Robson¹⁹ estimates a minimum of 5–6 km using only the throw of the major faults. Garfunkel and Bartov²⁰ suggest 15–20 km based on field work solely on the Sinai side of the Gulf. However, internal deformation of the blocks is widespread, complicating estimates of extension on geological grounds^{20,38}.

The uplift caused by extension of the lithosphere is a result of thermal expansion and is given by¹

$$U_L = \frac{A\alpha T_0}{2} (1 - 1/\beta_L)$$

where β_L is the average extension factor of the whole lithosphere.

To estimate the amount of extension required to create the observed uplift, a value for the lithosphere thickness must be assumed. The basement in the Gulf of Suez is of Pan-African age⁴² and thus had been quiescent for ~500 Myr before rifting. The northern coast of Egypt was formed by neo-Tethyan rifting in the Triassic, but the hinge zone of that event lies at least 200 km to the north of the profile studied, so that it probably did not greatly affect the thermal regime in the central Gulf of Suez. The compressional events of the Syrian Arc, however, do involve the northern part of the Gulf of Suez (ref 20 and B. Eggertson, personal communication). The extent to which the Syrian Arc events may have effected the thickness of the lithosphere is unknown. Thus, it is likely that the lithosphere thickness in the Gulf of Suez is close to 225 km, which was estimated by Karner *et al.*⁴³ to be the maximum thermal thickness of the continental lithosphere. This will result in a minimum value for

Table 1 Values used for analysis of extension

ρ_w	Density of water (g cm^{-3})	1.03
ρ_c	Density of crust (g cm^{-3})	2.8
ρ_m	Density of mantle (g cm^{-3})	3.33
α	Coefficient of thermal expansion ($^\circ\text{C}^{-1}$)	3.4×10^{-5}
T_0	Temperature at base of lithosphere ($^\circ\text{C}$)	1,333
C	Initial crustal thickness (km)	40–42
A	Initial lithosphere thickness (km)	125–225

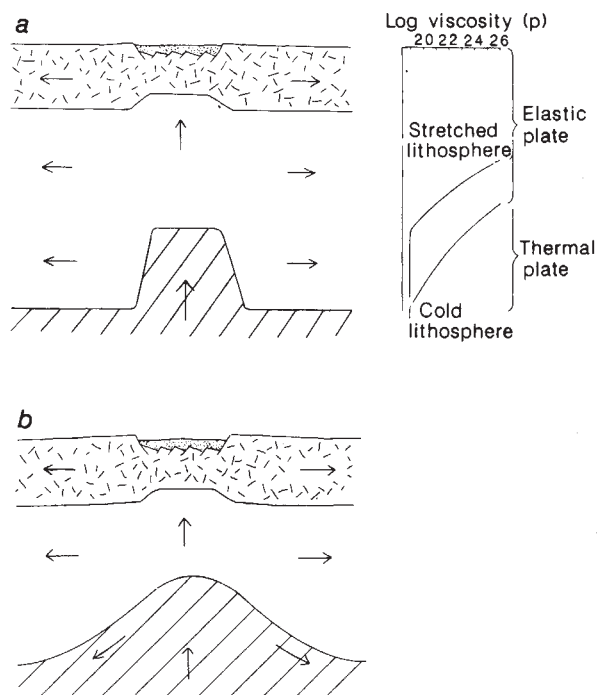


Fig. 4 *a*, Diagrammatic representation of geometry of a rift. Arrows indicate direction of flow implied by rifting; hatching shows rise of asthenosphere. The curve on the right-hand side shows variation in viscosity through the lithosphere for a power-law rheology with an activation energy of $100 \text{ kcal mol}^{-1}$ and a reference viscosity of 10^{19} . Rapid increase in viscosity will limit convection in the elastic plate and confine small-scale convection to the lower lithosphere. *b*, Diagram showing schematic geometry of rift resulting from small-scale convection.

the extension. The lower value used is 125 km. This is the asymptotic value for the oceans⁴⁴ and is the value generally assumed in extension models.

Figure 3 shows the distribution of extension based on Airy isostasy for these values and the total extension needed to produce the uplift. Inclusion of flexure would affect the distribution of extension, but not the total amount. The minimum extension of the total thermal lithosphere is 62 km which is 2.5 times greater than the extension represented by crustal thinning. If the crust has only extended by 25–27 km then the subcrustal lithosphere must have extended by a much greater amount. Variation of the parameters and estimates used in all stages of the analysis introduce a maximum uncertainty to ± 2 –3 km in the estimate of crustal extension and somewhat more in the lithosphere extension. The large ratio of lithosphere to crustal extension needed to match the observations remains. If account had been taken of the duration of rifting and the cooling which must have occurred since the initiation of the Gulf of Suez, then the lithosphere heating would be even greater. This is incompatible with uniform extension throughout the lithosphere and, therefore, the possibility that the heating of the lithosphere represented by the uplift was produced by uniform extension¹ must be rejected.

Discussion

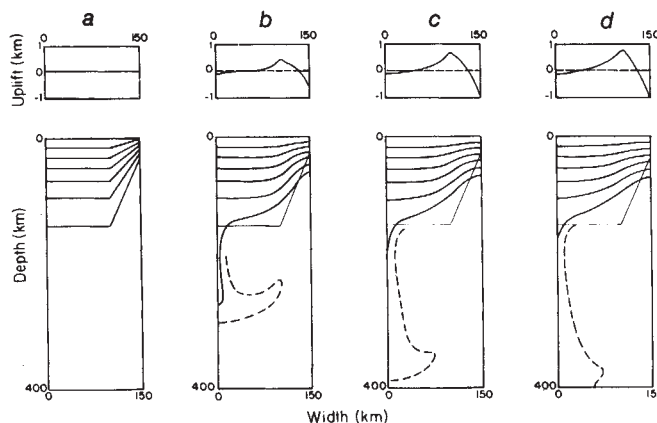
Several modifications to uniform extension have been proposed to account for differences between crustal and lithosphere extension estimates. Two-layered extension^{2,3} where the width of the extension zone varies with depth does not introduce excess heat and recent results by Keen⁴⁵ indicate that side driven extension occurs uniformly. Two-layer extension^{2,3} where the total extension of the layers is not equal produces strain incompatibility problems⁴⁶, so we must reject the possibility of actual multi-layer extension. It represents, at best, an *ad hoc* method of modelling the cooling history of an initial thermal state of unknown origins. Melt differentiation⁴ and underplating⁴⁷ could produce a discrepancy between lithosphere heating and crustal thinning. These mechanisms, which would thicken the crust during extension, imply that the crustal thinning underestimates the true extension and that the uplift should produce a truer estimate of actual extension. As the rift is only 90 km wide and entirely underlain by continental basement, 62 km is a physically implausible value for extension. The assumption of a thinner lithosphere yields even larger values, 112 km for a 125-km thick-

ness, which is greater than the width of the rift. Therefore, these models are untenable in the Gulf of Suez. I conclude that the estimate of extension based on crustal attenuation is correct and the amount of heating is at least 150% greater than explicable by extension.

Active rifting mechanisms in which a deep-seated source drives rifting is not compatible with the geological evidence. As discussed earlier, uplift did not precede rifting. The main phase of uplift did not develop until the Middle Miocene at least 8–10 Myr after the initiation of rifting. Fission track ages from the Sinai⁴⁸ indicate that 3 km (including erosion and isostatic rebound) of the uplift has occurred in the past 9 Myr. Major uplift continued despite a waning or cessation of rifting. This suggests that the extra heating was introduced by an intrinsic part of the rifting process.

The most likely mechanism of introducing extra heat as a result of rifting is through convective flow induced by the large horizontal temperature gradients developed by rifting. The geometry of a rift, with upwards and outwards flow of material and the juxtaposition of hot, buoyant asthenosphere and colder, denser lithosphere (Fig. 4) will greatly enhance convective flow within the mantle. Models that consider the lithosphere–asthenosphere boundary as a Rayleigh–Taylor instability^{12,13} usually also consider the entire mantle lithosphere as a single viscosity fluid and, therefore, conclude diapiric instabilities rapidly grow to the Moho. Fleitout and Froidevaux⁴⁹ examine the moment created by density anomalies in extension or compression and find that the lithosphere thermal anomaly is destabilizing, that is, acts to enhance the tectonics, and generally dominates over the stabilizing effect of the crustal thickness change. They also calculate flow in the lithosphere assuming a single viscosity for the mantle. In fact, the tremendous increase in viscosity with decreasing depth through the mantle will impede the development of flow in the upper part of the lithosphere (Fig. 4a). Full convective solutions which include temperature and pressure dependence of the viscosity^{50–52}, find that advection does not penetrate the upper lithosphere in a sharp uplift but broadens the width of the thermally perturbed zone, concentrating excess heat brought into the lithosphere in the rift flanks rather than beneath the rift. The strong mechanical lithosphere thickness responsible for the flexural strength of the lithosphere stabilizes the upper part of the lithosphere to flow. The lower part of the lithosphere, however, is effected by convection (Fig. 4b).

Fig. 5 Development of small-scale convection beneath rifts through time from numerical calculations in a fluid with pressure-dependent rheology (after Buck⁵⁰). Half width of rift is 50 km and is assumed to be symmetrical. The average β for the rift is 1.6, and the internal temperature of the box is 1,300 °C. Lower boxes show isotherms within lithosphere at 200 °C intervals. Dashed line, 1,250 °C isotherm, included to show convective flow. Light solid line, position of initial 1,200 °C isotherm. Comparison of 1,200 °C isotherms shows the sweeping away of material beneath flanks of rift producing broad uplift. Upper boxes show topography developed from convection and conduction. *a*, 0 Myr; *b*, 5 Myr; *c*, 10 Myr; *d*, 15 Myr.



In essence, the difference in thickness between the mechanical and thermal lithosphere leaves a region that contributes to the thermal isostasy and, therefore, vertical motion expressed at the surface of the Earth, but has little intrinsic strength to resist deformation or flow. In the oceans, this is the region termed the thermal boundary layer⁵³, which is susceptible to small-scale convection.

Numerical convection experiments in which a rift has been incorporated into the initial temperature structure clearly show the development of small-scale convection beneath the rift⁵⁰⁻⁵². Figure 5 shows the results of a numerical calculation by Buck⁵⁰ which includes a temperature-dependent viscosity. The relief on the isotherms greatly enhances the development of small-scale convection. The effect of this convective flow (Fig. 5) is to sweep away the 'corners' of lithosphere bordering the rift, thus creating broad uplifted flanks similar to those observed at the Gulf of Suez. The replacement of lithosphere by asthenosphere is concentrated in the thermal boundary layer. Including stress and pressure dependence should enhance the effect.

Although the models of Buck^{50,52} and Fleitout⁵¹ do not yet incorporate extension concurrent with convection, so the development of the uplift cannot be compared directly with the numerical results, it is encouraging that the wavelengths match those observed at the Gulf of Suez and that the amplitude and timing show good correlation.

This mechanism should be active beneath all rifts and continental margin basins. Karner and Weissel⁵⁴ suggest this mechanism explains the uplift of the south east Australian highlands. It can explain the widespread occurrence of rift flank

uplifts where there is no indication of active sub-lithosphere heat sources^{8,14}. There has also been a need for modelling using the *ad hoc* assumption of two-layer extension landwards of the hinge zone in continental margins^{2,4,55-57}. This can now be understood as a dynamic consequence of extension and reflects the extra heating that would occur at the landward edge of a continental margin caused by induced convection.

In summary, geological constraints on the timing of rift flank uplifts and basin analysis reconstruction of the crustal extension and lithosphere heating, when combined, indicate that the rift flank topography developed concurrently with major extension in the Gulf of Suez. Moreover, this uplift is caused by lithosphere heating greatly in excess of the thermal anomaly introduced by the 25-27 km of extension in the central Gulf of Suez. The most likely explanation for these observations is that thermal and velocity perturbations introduced by rifting induce secondary small-scale convection beneath the rift. The convection primarily heats the rift flanks, producing the broad uplift, and can explain heating in excess of extension at many rifts and basins.

I thank Barry Parsons for introducing me to the problem of border uplifts at the Gulf of Suez, AMOCO and GUPCO for support while in Egypt, and AMOCO, GUPCO, BP and EGPC for their generous release of data and valuable discussions. Roger Buck contributed greatly to my understanding of convection and allowed the use of his results in Fig. 5. Douglas Robson and Zvi Garfunkel helped improve my understanding of the geology of the gulf. J. Weissel, R. Buck and L. Fleitout provided helpful reviews. This research was supported by NSF grant OCE-83-09983. L-DGO contribution 3864.

Received 15 March; accepted 17 July 1985.

- McKenzie, D. P. *Earth planet. Sci. Lett.* **40**, 25-32 (1978).
- Royden, L. & Keen, C. E. *Earth planet. Sci. Lett.* **51**, 343-361 (1980).
- Slater, J. G. et al. *Earth planet. Sci. Lett.* **51**, 139-162 (1980).
- Beaumont, C., Keen, C. E. & Bouillier, R. *Geophys. J. R. astr. Soc.* (1982).
- Foucher, J. P., LePichon, X. & Sibuet, J. C. *Phil. Trans. R. Soc. A* **305**, 27-43 (1982).
- Morgan, P. *Tectonophysics* **94**, 277-298 (1983).
- Baker, B. H., Mohr, P. A. & Williams, L. A. *J. Bull. geol. Soc. Am. Spec. Pap.* **1136** (1972).
- Golombek, M. D., McGill, G. E. & Brown, L. *Tectonophysics* **94**, 483-507 (1983).
- Illies, J. H. & Greiner, G. *Bull. Geol. Soc. Am.* **89**, 770-782 (1982).
- Zorin, Y. A. *Tectonophysics* **73**, 91-104 (1981).
- Spohn, T. & Schubert, G. *J. geophys. Res.* **87**, 4669-4696 (1982).
- Neugebauer, H. J. *Tectonophysics* **94**, 91-108 (1983).
- Mareschal, J. C. *Tectonophysics* **94**, 51-66 (1983).
- Mohr, P. in *Geodynamics Series* Vol. 8 (ed. Palmason, G.) 293-309 (American Geological Union, Washington, DC, 1982).
- Sengör, A. M. C. & Burke, K. *Geophys. Res. Lett.* **5**, 419-421 (1978).
- Burke, K. & Whiteman, A. J. in *Implications of Continental Drift to the Earth Sciences* (eds Tarling, P. H. & Runcorn, S. K.) 735-755 (Academic, New York, 1973).
- Gass, I. G. *Phil. Trans. R. Soc. A* **267**, 369-381 (1970).
- Kent, P. *Tectonophysics* **36**, 87-92 (1976).
- Robson, D. A. Q. *J. geol. Soc. Lond.* **115**, 247-276 (1971).
- Garfunkel, Z. & Bartov, Y. *Bull. geol. Surv. Israel* **71**, 1-44 (1977).
- Scott, R. W. & Govean, F. M. *Paleogeogr. Paleoclimatol. Paleocool.* (in the press).
- Freund, R., Zak, I. & Garfunkel, Z. *Nature* **220**, 253-255 (1968).
- Freund, R. et al. *Phil. Trans. R. Soc. A* **267**, 107-130 (1970).
- Freund, R. *Nature* **228**, 453 (1970).
- Barakat, H. & Miller, P. M. *7th EGPC Explor. Sem.*, Cairo (EGPC, 1984).
- Tamsett, D. *Tectonophysics* **104**, 35-46 (1984).
- Martinez, F., Cochran, J. R., Hobart, M. A., Steckler, M. S. & Dehgheni, A. *Cont. Exten. Tect. Meet.*, Durham Abstr. (1985).
- Egyptian General Petroleum Company, stratigraphic committee *Oligocene and Miocene Rock-Stratigraphy of the Gulf of Suez Region* (EGPC, 1964).
- Zahrán, M. E. *GUPCO Explor. Dept. Rep.* ER-83-19 (1983).
- Beleily, A. *6th EGPC Explor. Sem.*, Cairo (EGPC, 1982).
- Moustafa, A. M. *5th EGPC Explor. Sem.*, Cairo (EGPC, 1976).
- Bunter, M. A. G. *6th EGPC Explor. Sem.*, Cairo (EGPC, 1982).
- Steen, G. E. *GUPCO Explor. Dept. Rep.* ER 82-1 (1982).
- Stecker, M. S. & Watts, A. B. *Earth planet. Sci. Lett.* **41**, 1-13 (1978).
- Ginzburg, A. et al. *J. geophys. Res.* **84**, 1569-1582 (1979).
- Ginzburg, A. & Gvirtzman, G. *Sedim. Geol.* **23**, 19-36 (1979).
- Healy, J. H. et al. *US Geol. Surv. Open File Rep.* OF-02-37 (1982).
- Chenet, P. Y. & Letouzey, J. *Bull. Centres Rech. Explor. Prof. Elf-Aquitaine* **7**, 201-215 (1983).
- Sultan, N. & Schütz, K. *7th EGPC Explor. Sem.* Cairo (EGPC, 1984).
- Cochran, J. R. *J. geophys. Res.* **86**, 263-288 (1981).
- Cochran, J. R. *Bull. Am. Ass. Petrol. Geol.* **67**, 41-69 (1983).
- Schurmann, H. M. E. *The Precambrian along the Gulf of Suez and the Northern Part of the Red Sea* (Brill, Leiden, 1965).
- Karner, G. D., Steckler, M. S. & Thorne, J. A. *Nature* **304**, 250-253 (1983).
- Parsons, B. & Slater, J. G. *J. geophys. Res.* **82**, 803-827 (1977).
- Keen, C. E. *Geophys. J. R. astr. Soc.* **80**, 95-120 (1985).
- Kligfield, R., Crespi, J., Naruk, S. & Davis, G. I. H. *Tectonics* **3**, 577-609 (1984).
- McKenzie, D. P. *Nature* **307**, 616-619 (1984).
- Kohn, B. P. & Eyal, M. *Earth planet. Sci. Lett.* **52**, 129-141 (1981).
- Fleitout, L. & Froidevaux, C. *Tectonics* **1**, 21-56 (1982).
- Buck, W. R. *Thesis*, Massachusetts Inst. Technol. (1984).
- Fleitout, L. thesis, Univ. Orsay (1984).
- Buck, W. R. *Earth planet. Sci. Lett.* (submitted).
- Parsons, B. & McKenzie, D. P. *J. geophys. Res.* **83**, 4485-4496 (1978).
- Karner, G. D. & Weissel, J. K. *7th Aust. geol. Conv.*, Abstr., 293 (1984).
- Steckler, M. S. *Thesis*, Univ. Columbia (1981).
- Watts, A. B. & Thorne, J. A. *Mar. Petr. Geol.* **1**, 319-339 (1984).
- Hellinger, S. J. & Slater, J. G. *J. geophys. Res.* **88**, 8251-8270 (1983).

A molecular solution to the riddle of the giant panda's phylogeny

Stephen J. O'Brien^{*†}, William G. Nash^{*}, David E. Wildt^{*†},
Mitchell E. Bush[†] & Raoul E. Benveniste^{*}

^{*} Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701, USA

[†] Department of Animal Health, National Zoological Park, Smithsonian Institution, Washington, DC 20008, USA

*Although it is generally agreed that the giant panda (*Ailuropoda melanoleuca*) is a member of the order Carnivora, there has long been disagreement over whether it should be classified with bears, raccoons or as a single member of its own family. Four independent molecular and genetic measures lead to a consensus phylogeny for the giant and lesser pandas. The lesser panda diverged from New World procyonids at approximately the same time as their departure from ursids, while ancestors of the giant panda split from the ursid lineage much later, just before the radiation which led to modern bears. The giant panda's divergence was accompanied by a chromosomal reorganization which can be partially reconstructed from the ursid karyotype, but not from that of procyonids or the lesser panda. The apparently dramatic, but actually limited, distinctions between the giant panda and the bears in chromosomal and anatomical morphology provide a graphic mammalian example of the discordance of molecular and morphological (and chromosomal) evolutionary change.*

THE taxonomic status of the giant panda and its diminutive Chinese relative, the lesser panda, has been a biological puzzle since their description by western naturalists a century ago^{1,2}. Although there is some agreement that the lesser panda was a member of the raccoon family (Procyonidae), a survey of over 40 published phylogenetic treatises³⁻⁵ shows that the giant panda has been classified with almost equal frequency in Ursidae (bear family), in Procyonidae and as the single member of a separate family, Ailuropodidae. The panda does look like a bear but has some characteristics and habits that are unusual for bears: (1) Despite its classification in the order Carnivora, it is largely herbivorous, subsisting primarily on bamboo. (2) The giant panda has extremely large forequarters and reduced hindquarters, which account for its ambling gait. The enlargement of the anterior end is associated with a huge specialized masticatory apparatus which has some parallel in the lesser panda. (3) The male genitalia in the giant panda are tiny, cylindrical, S-shaped and posteriorly directed, very similar to the lesser panda. (4) The giant panda has six digits on its forepaw, resulting from the evolutionary extension of the radial sesamoid to an awkward, but functional, opposable thumb^{3,6}. (5) Finally, the giant panda does not behave like a bear^{7,8}. Most bears hibernate, the giant panda does not; bears roar, whereas the giant panda bleats^{7,8}.

The proponents of grouping of the giant and lesser panda in the Procyonidae have emphasized the similarity of the two in tooth structure, skull architecture, penis morphology and colour patterning^{4,9}. Davis³, using anatomical data, and Leone and Wiens¹⁰ and Sarich¹¹, using serological data, classified the panda as a bear. Davis's conclusions were disputed as being uncritical^{12,13} and the immunological data were often ignored¹⁴. Moreover, recent fossil evidence described by Chinese palaeontologists¹⁵⁻¹⁷ and the widely divergent karyotype of bears versus pandas^{18,19} lent support to the separate family interpretation.

Here we report the use of various molecular methods to estimate evolutionary distances between the giant panda, the lesser panda and their supposed closest relatives. We considered three molecular techniques: (1) DNA hybridization between species using unique-sequence cellular DNA; (2) genetic distance based on electrophoretic mobility of >50 homologous isozyme loci; and (3) immunological distance of serum proteins. Analysis and interpretation of quantitative differences involving these techniques in extant species is based on the 'molecular clock' hypothesis²⁰⁻²². As a control for the procedures and as a strategy for calibrating evolutionary time in these species, we performed the same estimates on tissues of the Hominoidea (apes and man), an extensively studied group²³⁻²⁹. Finally, a comparative cytological analysis of highly extended G-banded

chromosomes from pandas and bears is presented. Our results converge on a consensus phylogeny of the pandas.

DNA hybridization

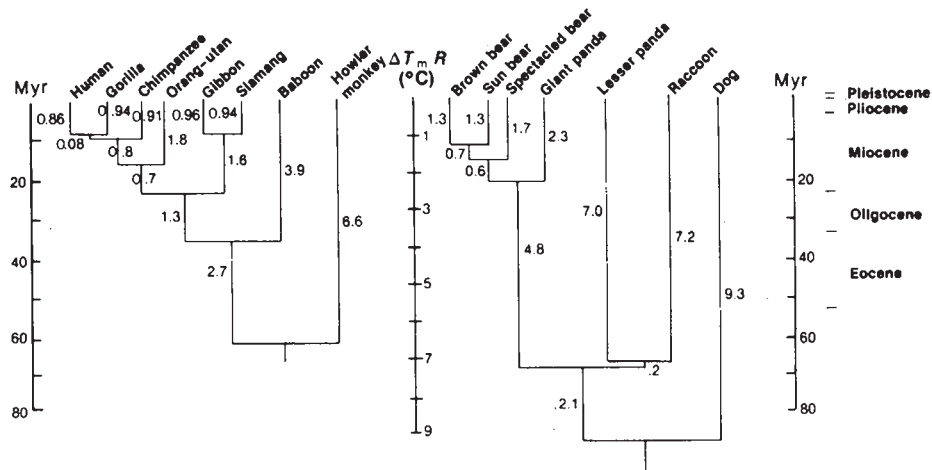
DNA hybridization of unique-sequence cellular DNA between heterologous species is a common technique in the measurement of phylogenetic affinity²⁸⁻³². Primary fibroblast cells from four primate index species (*Homo sapiens*, human; *Pan troglodytes*, chimpanzee; *Gorilla gorilla*, gorilla; and *Hylobates concolor*, black gibbon) were cultured in the presence of radiolabelled ³H-thymidine and their high C₀t (≥200) DNA was purified using hydroxyapatite. This presumed 'single-copy' DNA was hybridized, using an S₁ nuclease assay²⁹, to excess cold DNA extracted from the same species and from three additional primates (Fig. 1). Similarly, reciprocal DNA hybridization curves were derived between each of four index carnivore species (*Procyon lotor*, raccoon; *Ailuropoda melanoleuca*, giant panda; *Ailurus fulgens*, lesser panda; and *Ursus arctos*, American brown bear). Changes in the average thermal stability (see Fig. 1 legend) were also measured between these species and two additional ursids: *Tremarctos ornatus*, spectacled bear; and *Helarctos malayanus*, Malayan sun bear (Fig. 1).

Data matrices (Fig. 1) were analysed using five distinct phylogenetic algorithms: the distance Wagner procedure³³⁻³⁵, the UPGM algorithm³⁶, the neighbourliness method³⁷, the MATTOP program³⁸ and the Fitch-Margoliash³⁹ algorithm. All the algorithms produced phylogenetic trees that support the presented topologies, although the actual distances varied depending on the method used. The hominid tree shows an excellent agreement both in topology and in arm lengths with trees developed elsewhere²⁵⁻²⁸, the only confusion being the trichotomy between African apes and man which has been under debate for some time^{26-29,31,40-43}. The derived carnivore topology places the giant panda as an early leg of the ursid lineage occurring before the divergence of the modern bears but after the ursid-procyonid split. The first bear in the ursid proper radiation was the ancestor of the spectacled bear (*T. ornatus*) which conforms to its morphological assignment to generic status in the family Ursidae⁴⁴. The lesser panda, the ursids and the procyonids seemed to split from each other at approximately the same time. Although the tree (Fig. 1) favours an association between the raccoon and the lesser panda, the errors associated with the large distances between lesser panda, raccoon and the bears do not allow unequivocal resolution of the trichotomy.

Isozyme genetic distance

Isozyme genetic distance is a statistical calculation, developed

Fig. 1 Average thermal stabilities ($\Delta T_m R$, °C; 2-9 experiments) derived from DNA hybridization of genomic DNAs of indicated species. Primary fibroblast cultures were established for each species from skin or organ biopsies of primate and carnivore species. Cultured cells from four primates (human, chimpanzee, gorilla and gibbon) and four carnivores (brown bear, giant panda, lesser panda and raccoon) were grown in the presence of ^3H -thymidine and the non-repetitive DNA was isolated and hybridized to each of the species listed in the left-hand column of each table²⁹. ΔT_m is the difference in T_m between the other DNA-DNA hybrids and the T_m of the homologous hybrids. $\Delta T_m R$ is the thermal stability difference for all non-repeated DNA, and is the ΔT_m value corrected for the normalized percentage of hybridization to the labelled index species²⁸⁻³⁰. Distance matrices for phylogenetic inference were judged to be adequate, based on four different measurements: first, the standard deviations of the estimates (which were based on 2-9 hybridizations per pair) were low ($\pm 0.2^\circ\text{C}$ for $\Delta T_m R < 5^\circ\text{C}$, and $\pm 0.5^\circ\text{C}$ for $\Delta T_m R$ values $\geq 5^\circ\text{C}$). Second, the reciprocal estimates were in close agreement, as evaluated by calculating the mean % deviation from the mean⁶⁰: 2.1% for the primates and 1.8% for carnivores. Third, the metricity of the distance matrix is evidenced by conformance of all the three-way combinations to the triangle inequality^{33,35}. Metricity or conformance to the triangle inequality is not an absolute indicator of clock-like behaviour because a distance may be metric but still exhibit variation in evolutionary rates³⁵. Thus, metricity is a necessary, but not sufficient, condition for an evolutionary clock. Fourth, the apparent constancy of the DNA clock was evaluated by demonstrating that internal species were equidistant from an outside group (howler monkey and dog, respectively), the relative rate test^{21,40}. The phylogenetic trees shown here were derived from the Fitch-Margoliash algorithm³⁹ based on maximum parsimony. The actual trees were drawn to scale using the KITSCH of the PHYLIP program, given by J. Felsenstein (University of Washington). This program computes a rooted topology based on the assumption of an evolutionary molecular clock rendering all terminal species as contemporaneous. The numbers are the leg lengths of the unrooted tree generated by the Fitch-Margoliash algorithm in the absence of the above assumptions. This tree is rooted by the midpoint of the two most distal species in the network. Placing a timescale with some degree of confidence on the three derived molecular topologies is difficult. Although the carnivore fossil record is unusually good and has been studied intensively, even the commonly accepted dates for divergence of these families can be incorrect by up to 25-50%. For example, the best geological dates for the time of procyonid-ursid divergence vary from 30 to 50 Myr ago. A strategy we have used for setting our molecular clock was to take advantage of the demonstration that the primate and carnivore clocks seem to run at the same rate^{31,48}. We use a timescale based on Pilbeam's date of the human-orang-utan divergence occurring at ~16 Myr BP^{28,31,53}. Because the human-orang-utan split occurred near the middle of the timescale, the errors are not amplified at the stem or the root of the trees. The derived topology places the hominid-Old World monkey split at 35 Myr BP, a date which is not at variance with fossil or molecular conclusions previously presented^{28,29,40,53,61}. NT, not tested.

Thermal stabilities ($\Delta T_m R$, °C)Thermal stabilities ($\Delta T_m R$, °C)

	Human	Chimpanzee	Gorilla	Gibbon		Brown bear	Giant panda	Lesser panda	Raccoon
Human	—	1.9	1.8	5.2					
Chimpanzee	1.8	—	1.9	4.9	Brown bear	—	4.3	14.0	14.4
Gorilla	1.8	2.0	—	5.0	Sun bear	2.5	NT	NT	NT
Orang-utan	3.5	NT	3.6	NT	Spectacled bear	3.3	NT	NT	NT
Gibbon	5.0	5.0	5.1	—	Giant panda	4.8	—	14.4	14.2
Siamang	5.1	4.9	5.0	1.9	Lesser panda	13.9	14.1	—	14.3
Baboon	7.7	7.6	7.8	7.7	Raccoon	14.4	14.7	13.9	—
Howler	13.0	13.3	NT	13.1	Dog	18.7	18.3	18.9	18.5

and improved by Nei^{45,46}, for measuring the degree of allelic substitutions at a group of loci between populations (or species) based on the electrophoretic mobility of soluble proteins. The distance estimate, D , is defined as the average number of gene differences per locus between individuals from two test populations. Within the limits of certain specific assumptions relating to electrophoretic resolution and relative rates of nucleotide substitution, the genetic distance estimates increase proportionately with the amount of time the compared populations have been reproductively isolated^{43,45-47}. The derived ape topology (Fig. 2) agrees with the DNA topology (Fig. 1) and with previous studies in placing the orang-utan as an early branch before the unresolved trichotomy between man and African great apes.

The same five phylogenetic algorithms applied to the DNA data were used to derive a topology of the giant panda and its relatives using isozyme data (Fig. 2). In general, the most parsimonious and most consistently derived relationship was the same and differed only in branch length. As illustrated by the Fitch-Margoliash tree (Fig. 2), the giant panda diverged from the ursid line just before the radiation of modern bears, but long after the ursid-procyonid split. After the giant panda divergence, the next split involved the spectacled bear, followed by a three-way split (another unresolved trichotomy) between the brown, black and sun bears. The lesser panda and the raccoon seem to have shared a common ancestor for a short time after procyonid divergence.

The same five phylogenetic algorithms applied to the DNA data were used to derive a topology of the giant panda and its relatives using isozyme data (Fig. 2). In general, the most parsimonious and most consistently derived relationship was the same and differed only in branch length. As illustrated by the Fitch-Margoliash tree (Fig. 2), the giant panda diverged from the ursid line just before the radiation of modern bears, but long after the ursid-procyonid split. After the giant panda divergence, the next split involved the spectacled bear, followed by a three-way split (another unresolved trichotomy) between the brown, black and sun bears. The lesser panda and the raccoon seem to have shared a common ancestor for a short time after procyonid divergence.

Immunological distance

The third method is that of immunological distance^{21,22} and we summarize here the results of Sarich¹¹. The procedure measures the displacement of immunological titration curves between

species based on amino-acid substitutions that have occurred in homologous proteins. The extensive and comprehensive molecular evolutionary studies of Sarich, Wilson and collaborators have demonstrated the constancy of the albumin clock in several vertebrate taxa by using microcomplement fixation as a measure of immunological distance^{11,21,22,40,48,49}. Sarich prepared both albumin and transferrin (from two species, brown bear and raccoon) to generate a carnivore tree which is reproduced in Fig. 3. These results also placed the giant panda on the ursid line before the divergence of modern bears, but after ursid-procyonid split. The position of the lesser panda gave conflicting results with albumin (which favoured the ursid lineage) versus transferrin (which placed the lesser panda equidistant from both bear and raccoon), probably because of the error associated with the large distances involved. Although he shows caution in interpreting the lesser panda placement, Sarich's main conclusion was clearly supported by both albumin and transferrin immunological distances.

Karyological evidence

The karyological relationship between the bears, panda and raccoon seems to emphasize the differences between bears and giant pandas, but, on closer analysis, has become particularly informative. The giant panda has 42 metacentric chromosomes whereas members of the genus *Ursus* (black, Malayan sun, and brown bears) have 74 mainly acrocentric chromosomes^{18,19}. Wurster-Hill and Bush¹⁹ prepared G-trypsin-banded karyotypes of the giant panda and were able to recognize only three homologies between giant panda and bear chromosomes. To extend these observations to a higher level of resolution, we transformed brown bear and giant panda primary fibroblasts

Fig. 2 Isozyme genetic distance (D) computed from 44 loci in primates and 50 loci in carnivores. The enzymes studied are homologous with enzymes examined previously in the domestic cat and man⁶²⁻⁶⁵.

Electrophoretic procedures were predominantly on starch gels using standardized protocols. For the primates, 44 enzyme systems were resolved: ACP1, -2, ADA, APRT, AK, ALB, ACY, CAT, CPKB, DIA4, ENO1, ES1, ESD, GALA, GDH1, GAPD, GLO1, GOT1, 2, G6PD, GPI, GSR, GUS, HBB, HK1, IDH1, -2, LDHA, -B, MDH1, MPI, PEPA, -B, PGM1, -2, -3, 6-PGD, PGAM, NP, PP, PK, SOD1, -2 and TPI (see ref. 63 for nomenclature). The same systems, with the exception of GAPD, were scored in the carnivores. In addition, the following enzymes were resolved in the carnivore study: FUCA, GPT, HEXA, PEPC, ES2 and ES3⁶³. Each system

was examined in extracts of erythrocytes, leukocytes and cultured fibroblasts for each species. The single exception was the American black bear, for which only cultured cells were available. Hence, only 36 of the 50 test loci could be evaluated. This difference dramatically affected the data, as the measured distances to other species were consistently greater than expected from the phylogenetic relationship. Because of the inequivalence of D values for the black bear, these data were not considered in the computation of outside-group distances. Genetic distances were calculated using the methods of Nei^{45,46}. Phylogenetic trees are derived using the Fitch-Margoliash algorithm, as described in Fig. 1 legend. The other algorithms³³⁻³⁹ produced similar trees with these data. Fitch's unique neighbourliness algorithm³⁷ yielded nearest-neighbour values which showed that the lesser panda and raccoon had the highest scores, that is, they are the nearest neighbours most often in four-way comparisons, followed closely by the three ursids sun, brown and black bears.

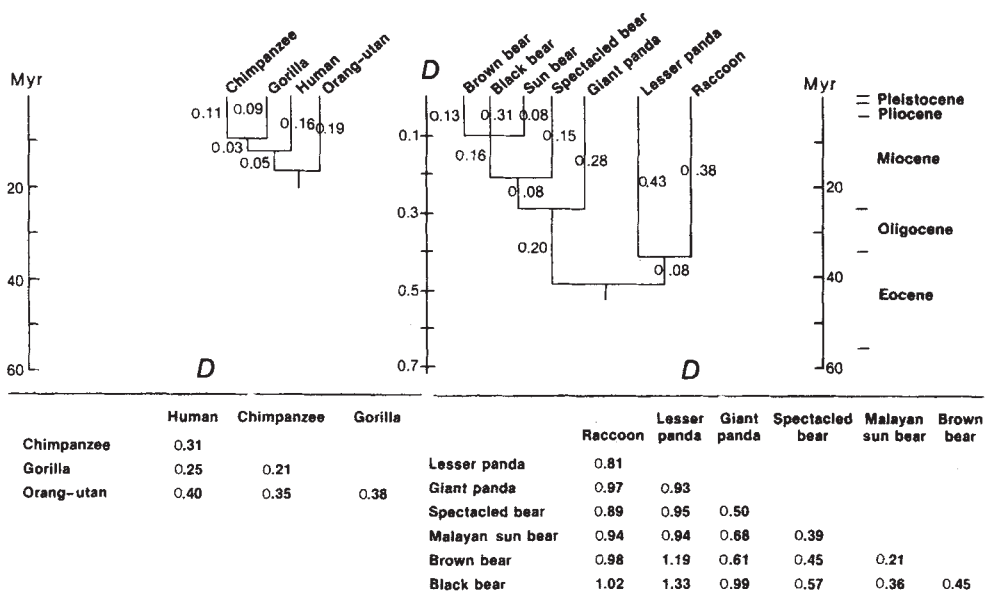
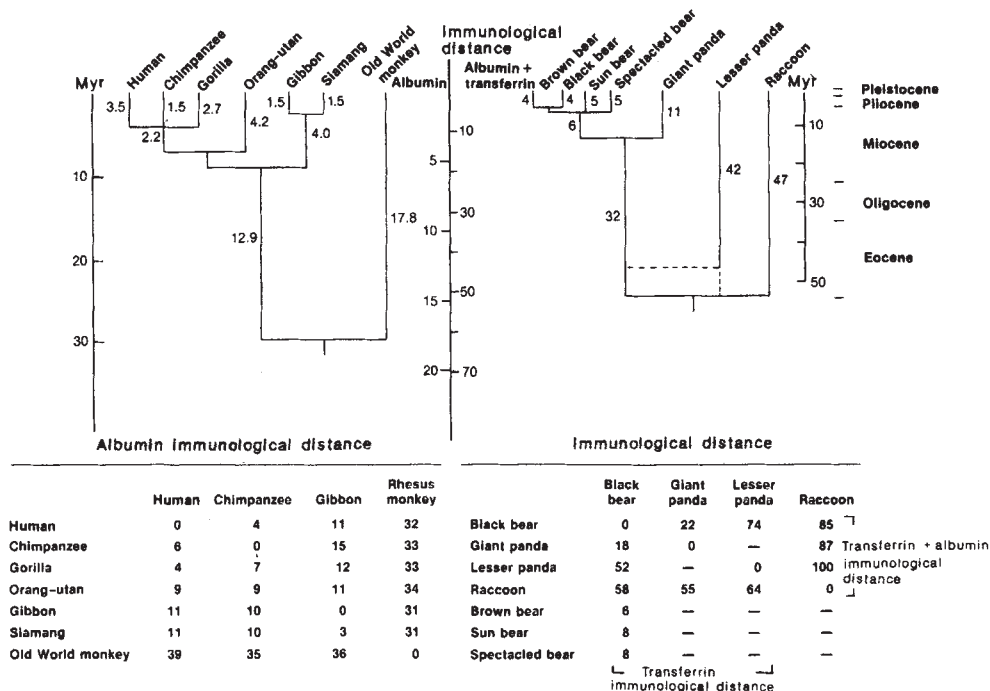


Fig. 3 Phylogenetic relationships of primates and carnivores based on immunological distance. Immunological distance was defined as $100 \times$ the logarithm of the index of dissimilarity, the factor by which the dilution of a heterologous serum is increased to fix the amount of complement fixed in the homologous reaction. Primate distances are derived from the albumin index of dissimilarity presented by Sarich and Wilson⁴⁰. Carnivore distances were based on transferrin plus albumin immunological distance data (upper right of table)¹¹. The topology was constructed using the Fitch-Margoliash procedure³⁹ outlined in Fig. 1. The additional ursid species were added to the tree using the transferrin immunological distances presented in the lower left of the table. Because of the inequivalence of evolutionary rates of the three immunological distance metrics, the alignment of the two trees was determined on the basis of the timescale presented by Sarich and Wilson⁴⁰ and by Sarich¹¹; this timescale is shown on the left of the figure, while the Pilbeam⁵³ timescale is presented on the right.



with an oncogenic retrovirus (feline sarcoma virus) and examined the preparations of extended early-metaphase chromosomes by G-trypsin banding (Fig. 4). Remarkably, nearly every large chromosome of the brown bear could be aligned with a giant panda chromosome arm. Sixteen such homologous comparisons are presented in Fig. 4. Alignments of smaller bear chromosomes (for example, UAR 19-24 in Fig. 4) were not attempted because homologies of 1-3 bands are likely to be coincidental. Nonetheless, the giant panda chromosomes seemed to be composed largely of bear chromosomes fused together as robertsonian translocations.

Only two of the banded giant panda or bear chromosomes had recognizable counterparts in the lesser panda or raccoon, thereby emphasizing the chromosomal consanguinity between bears and the giant panda¹⁹. On the contrary, 14 chromosomes of the lesser panda were strikingly homologous to chromosomes found in several procyonids and 10 of these were recognized as conservative ancestral 'carnivore chromosomes' found in several carnivore families other than Ursidae⁵⁰⁻⁵². The simplest explanation of these data seems to affirm the molecular topologies; that is, the lesser panda and procyonids share a common ancestor at or subsequent to ursid divergence. The ursid line apparently

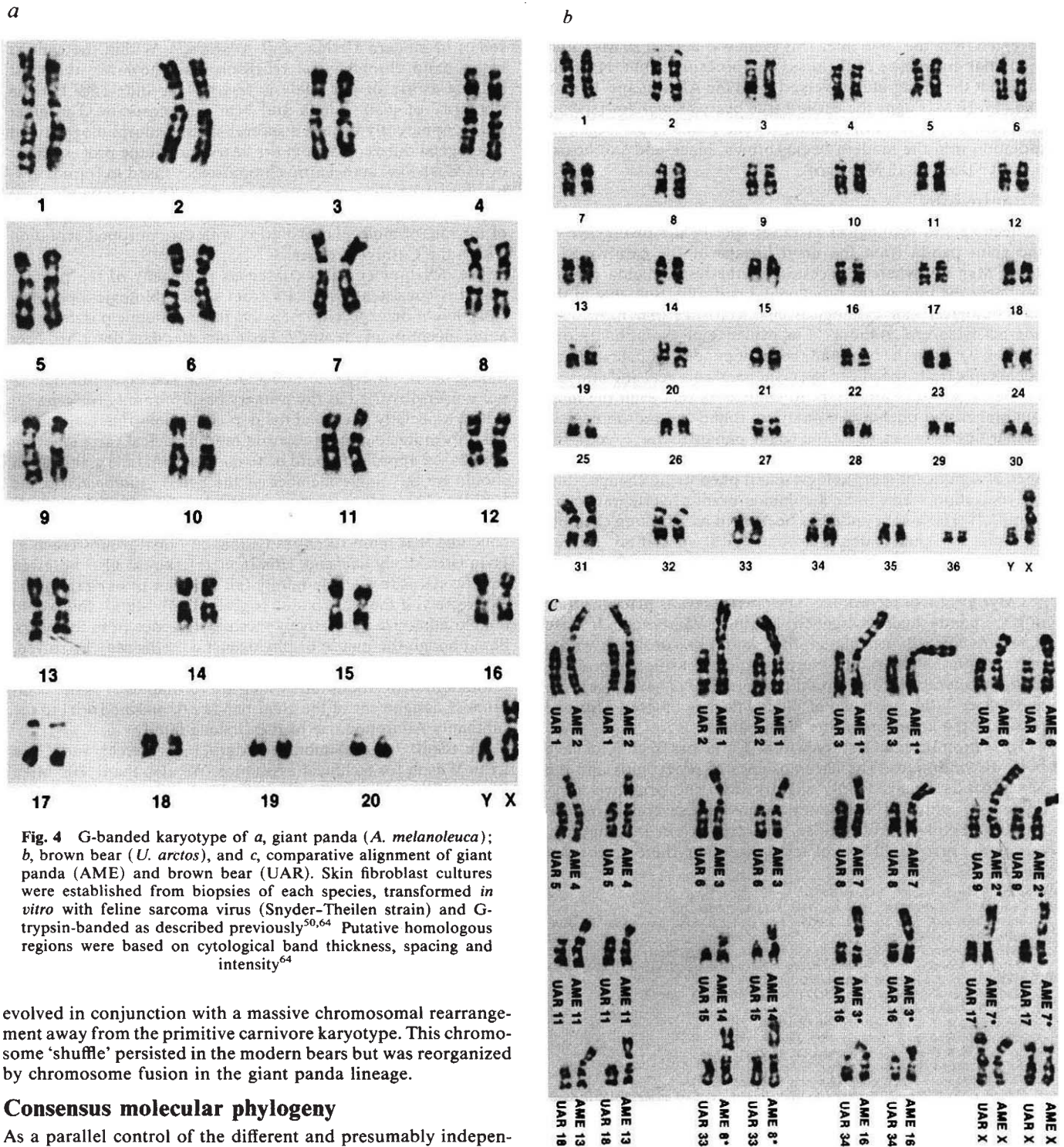


Fig. 4 G-banded karyotype of *a*, giant panda (*A. melanoleuca*); *b*, brown bear (*U. arctos*), and *c*, comparative alignment of giant panda (AME) and brown bear (UAR). Skin fibroblast cultures were established from biopsies of each species, transformed *in vitro* with feline sarcoma virus (Snyder-Theilen strain) and G-trypsin-banded as described previously^{50,64}. Putative homologous regions were based on cytological band thickness, spacing and intensity⁶⁴.

evolved in conjunction with a massive chromosomal rearrangement away from the primitive carnivore karyotype. This chromosome 'shuffle' persisted in the modern bears but was reorganized by chromosome fusion in the giant panda lineage.

Consensus molecular phylogeny

As a parallel control of the different and presumably independent techniques used here, we have applied the same methods to the construction of evolutionary trees of Hominoidae and representative simian species (Figs 1-3). Based on the 16-Myr calibration of the man-orang-utan split⁵³, the three primate phylogenies are topologically equivalent and in good agreement with the prevailing consensus of relationships between this group of primates^{23-28,40-42}. Basically, the apes diverged from their common ancestor with the Old World monkeys 30-40 Myr BP. The gibbons split next between 20 and 25 Myr ago. The great apes began their radiation ~13-16 Myr ago with the orang-utan lineage followed by the ape-gorilla-human split 8-10 Myr ago. The agreement of our DNA and isozyme data with the immunological distance data and other topologies emphasizes the reliability of application of these same procedures to the carnivores.

The three molecular methods were also highly consistent in

the derived relationship within the carnivore radiations. Furthermore, a comparison of the primate and carnivore topologies using each procedure allows the correlation of the bear radiations with primate events regardless of the calibration date(s) used. Briefly, between 30 and 50 Myr ago, the progenitors of modern ursids and procyonids split into two lineages. Within 10 Myr of that event (possibly at its inception), the procyonid group split into Old World procyonids (represented today by the genus *Ailurus*, the lesser panda) and the New World procyonids (for example, raccoon, coatis, olingo and kinkajou). The lesser panda and giant panda clearly do not share a common ancestor after the ursid-procyonid split, emphasizing that the morphological similarities of the pandas are probably the result of parallel retention of ancestral characters that may have been lost (for example, in the bear) after their divergence from the

main line. At about the same time as the gibbons split from great apes (18–25 Myr ago), the ancestor of the giant panda diverged from the ursid line. This event was at least 20 Myr after the initial divergence of the ursid and procyonid split. Near the time that the orang-utan diverged from the African ape–human line (13–16 Myr ago) the earliest true bear, *Tremarctos* (spectacled bear), split from the ursid line. The genus *Ursus* began its radiation into the modern bears (brown, black and sun bears) 6–8 Myr later (8–12 Myr ago).

Conclusions

Molecular and cytological methods specify the divergence of the giant panda from the ursid lineage of the carnivores at 15–25 Myr BP, whereas ancestors of the lesser panda emerge very near the time of the procyonid–ursid split. But what of the morphological and cytological characteristics that have been offered by several authors^{2,4,9} as evidence of evolutionary distinction between the bear and the giant panda? Of course, these phenotypic traits specific to the giant panda are real and unusual for bears, but it is important to emphasize that even the most comprehensive phenotypic analysis³ found the morphological similarities between bear and giant panda to far exceed the number (and extent) of the differences. Furthermore, the appearance of significant morphological and phenotypic changes during speciation seems to be a common event of vertebrate evolution^{54–56}. The giant panda has been cited as a striking example of rapid (or 'punctuated') morphological speciation⁵⁶, as its apparently dramatic anatomical divergence from bears (or raccoons) is accompanied by only a very recent (Pleistocene, <2 Myr BP) fossil record^{15–17}. The present results place the time of giant panda–bear divergence back to 15–25 Myr ago. If these dates are correctly calculated, they would be consistent with a more traditional or gradual morphological transition to produce modern *Ailuropoda*, and would also predict, as have Eisenberg and Setzer¹³, the existence of older Pliocene or even Miocene fossils of the ancestors of the giant panda.

The chromosomal history within Carnivora seems to have been discontinuous. The chromosomes of procyonids and the lesser panda changed only slightly from the primitive karyotype^{50–52}, which is today represented in several carnivore families (Felidae, Procyonidae, Viverridae). On the contrary, during the first 10–20 Myr of ursid evolution, there occurred a

quantum chromosomal reorganization that is reflected today by the almost complete absence of 'carnivore-specific' chromosomes in modern Ursidae and *Ailuropoda*. Within the context of an ursid chromosomal reorganization, however, the most striking aspect of our analysis was not the difference but the similarity of giant panda and bear chromosomes. The giant panda apparently acquired several specific morphological and ethological differences from the bears and, in the process, most of its 'bear-like' acrocentric chromosomes fused to form a low-numbered, largely metacentric karyotype. The giant panda–bear phylogeny seems to provide a specific example of the uncoupling of the rate of molecular evolution with chromosomal evolution within the Carnivore order^{57–59}.

The development of a consensus phylogeny of the pandas, which is consistent with data from several biological perspectives, might be expected to resolve the taxonomic puzzle posed at the inception of the study. The molecular data described here provide a clear relationship between the species and narrow the timing of the divergence nodes between the taxa. But time is not universally used as the primary basis of family (or generic) status, especially in view of the non-continuous timescale of the possibly more adaptively relevant morphological variation. One time-based approach would be to conclude that the giant panda should be the single member of the genus *Ailuropoda* in the Ursidae. The phylogenies presented here would not preclude this assignment; however, if we accept this, should we not also conclude that more recent radiations in other groups (such as Hylobates, Pongidae and Hominoidea) should also be given generic status in a single family (all the apes in our example)? Conversely, if the giant panda is given family status, this would tend to minimize its relatively recent divergence from the bears. So, although the puzzle of phylogenetic timing may be solved, classification of the pandas may continue to be argued. One resolution of this question (which we recommend) is the compromise assignment of the giant panda (*A. melanoleuca*) to the subfamily Ailuropodinae of the Ursidae family.

We thank Janice Simonson, Mary Eichelberger and Gaye Lynn Wilson for technical assistance. We also thank Drs Mike Braun, Elise Brownell, Glen Collier, Lisa Forman, Kurt Benirschke, George Johnson, Ross MacIntyre, Bill Modi, Alan Templeton and Doris Wurster-Hill for critical reading of early versions of the manuscript.

Received 19 March; accepted 24 June 1985.

- David, A. *Nouv. Arch. Mus. Hist. nat., Paris* 5, 12–18 (1869).
- Milne-Edwards, A. *Ann. Sci. nat. Zool. Ser. 5 art 10*, 1; *C.r. heb. Séanc. Acad. Sci., Paris* 70, 341–342 (1870).
- Davis, D. *Fieldiana, Zool. Mem.* 3 (Chicago Nat. Hist. Museum, 1964).
- Morris, D. & Morris, R. in *The Giant Panda*, 160–169 (Penguin, New York, 1981).
- Thienius, V. E. *Z. Säugetierk. Mitt.* 44, 286–305 (1979).
- Gould, S. J. in *The Panda's Thumb*, 19–34 (Norton, New York, 1980).
- Peters, G. *Z. Säugetierk. Mitt.* 47, 236–246 (1982).
- Kleiman, D. G. *Z. Tierpsychol.* 62, 1–46 (1983).
- Mivart, St G. *Proc. zool. Soc. Lond.* 1885, 340–404 (1885).
- Leone, C. A. & Wiens, A. L. *J. Mammal.* 37, 11–23 (1956).
- Sarich, V. *Nature* 245, 218–220 (1973).
- Ewer, R. *The Carnivores* (Cornell University Press, Ithaca, 1973).
- Collins, L. R. & Page, J. K. *Ling-Ling and Hsing-Hsing, Year of the Panda* (Anchor Press/Doubleday, Garden City, 1973).
- Gribbin, J. & Chertas, J. *The Monkey Puzzle; Reshaping the Evolutionary Tree* (Pantheon, New York, 1982).
- Wang, T. K. *Acta zool. sin.* 20, 191–201 (1974).
- Chu, C. *Acta zool. sin.* 20, 174–187 (1974).
- Pei, W.-C. *Acta zool. sin.* 20, 188–190 (1974).
- Newman, R. & Davidson, M. *Cytogenetics* 5, 152–163 (1966).
- Wurster-Hill, D. & Bush, M. *Cytogenet. Cell Genet.* 27, 147–154 (1980).
- Zuckerland, E. & Pauling, L. in *Horizons in Biochemistry* (eds Kasha, M. & Pullman, B.) 189–225 (Academic, New York, 1962).
- Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* 46, 573–639 (1977).
- Thorpe, J. P. *A. Rev. Ecol. Syst.* 13, 139–168 (1982).
- King, M. C. & Wilson, A. C. *Science* 188, 107 (1975).
- Kohne, D. E., Chiscon, J. A. & Hoyer, B. H. *J. hum. Evol.* 1, 627–644 (1972).
- Bruce, E. J. & Ayala, F. J. *Nature* 276, 264–265 (1978).
- Farris, S. D., Wilson, A. C. & Brown, W. M. *Proc. natn. Acad. Sci. U.S.A.* 78, 2432–2436 (1981).
- Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. *J. molec. Evol.* 18, 225–239 (1982).
- Sibley, C. G. & Ahlquist, J. E. *J. molec. Evol.* 20, 12–15 (1984).
- Benveniste, R. E. & Todaro, G. J. *Nature* 261, 101–108 (1976).
- Britten, R. J. & Kohne, D. E. *Science* 161, 529–540 (1968).
- Benveniste, R. in *Molecular Evolutionary Genetics* (ed. MacIntyre, R. J.) 359–417 (Plenum, New York, 1985).
- Brownell, E. *Evolution* 37, 1034–1051 (1983).
- Farris, J. S. in *Advances in Cladistics* (eds Funk, V. A. & Brooks, D. R.) (New York Botanical Garden, 1982).
- Felsenstein, J. *Evolution* 38, 16–24 (1984).
- Farris, J. S. *Cladistics* 1, 51 (1985).
- Sneath, P. H. A. & Sokal, R. R. in *Numerical Taxonomy*, 230–234 (Freeman, San Francisco, 1973).
- Fitch, W. M. *J. molec. Evol.* 18, 30–37 (1981).
- Dayhoff, M. O., *Atlas of Protein Sequence and Structure* 1–8 (National Biomedical Research Foundation, Washington, DC, 1976).
- Fitch, W. M. & Margoliash, E. *Science* 155, 279–284 (1967).
- Sarich, V. & Wilson, A. C. *Science* 158, 1200–1204 (1967).
- Templeton, A. R. *Evolution* 37, 221–224 (1983).
- Lewin, R. *Science* 226, 1179 (1984).
- Nei, M. *Molecular Population Genetics and Evolution* (North-Holland, Amsterdam, 1975).
- Nowak, R. M. & Paradiso, J. L. *Walker's Mammals of the World* 4th edn (Johns Hopkins University Press, Baltimore, 1983).
- Nei, M. *Am. Nat.* 106, 283–292 (1972).
- Nei, M. *Genetics* 89, 583–590 (1978).
- Avise, J. C. & Aquadro, C. F. in *Evolutionary Biology* (eds Hecht, M. K., Steere, W. C. & Wallace, B.) 151–185 (Plenum, New York, 1981).
- Sarich, V. *Syst. Zool.* 18, 286–295 (1969).
- Sarich, V. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* 58, 142–148 (1967).
- Wurster-Hill, D. H. & Gray, C. W. *Cytogenet. Cell Genet.* 12, 377–397 (1973).
- Wurster-Hill, D. H. & Gray, C. W. *Cytogenet. Cell Genet.* 15, 306–331 (1975).
- Wurster-Hill, D. H. & Centerwall, W. R. *Cytogenet. Cell Genet.* 34, 178–192 (1982).
- Pitbeam, D. *Nature* 295, 232–234 (1982).
- Gould, S. J. & Eldredge, N. *Paleobiology* 3, 115–151 (1977).
- Stanley, S. M. *Proc. natn. Acad. Sci. U.S.A.* 72, 646–650 (1975).
- Stanley, S. (ed.) *Macroevolution: Patterns and Processes* (Freeman, San Francisco, 1979).
- Wilson, A. C., Bush, G. L., Case, S. M. & King, M. C. *Proc. natn. Acad. Sci. U.S.A.* 12, 5061–5065 (1975).
- Bush, G. L., Case, S. M., Wilson, A. C. & Patton, J. L. *Proc. natn. Acad. Sci. U.S.A.* 74, 3942–3946 (1977).
- O'Brien, S. J., Seanez, H. N. & Womack, J. E. in *Molecular Evolutionary Genetics* (ed. MacIntyre, R. J.) 519–589 (Plenum, New York, 1985).
- Champion, A. B., Prager, E. M., Wachter, D. & Wilson, A. C. in *Biochemical and Immunological Taxonomy of Animals* (ed. Wright, C. A.) 397–416 (Academic, New York, 1974).
- Leakey, R. E. & Lewin, R. *Origins, the Emergence and Evolution of Our Species and Its Possible Future* (Dutton, New York, 1982).
- O'Brien, S. J. & Nash, W. G. *Science* 216, 257–265 (1982).
- Shows, T. & McAlpine, P. *Cytogenet. Cell Genet.* 37, 340–393 (1984).
- Nash, W. G. & O'Brien, S. J. *Proc. natn. Acad. Sci. U.S.A.* 79, 6631–6635 (1982).
- Harris, H. & Hopkinson, D. A. *Handbook of Enzyme Electrophoresis in Human Genetics* (North-Holland, Amsterdam, 1976).

Structure of a human common cold virus and functional relationship to other picornaviruses

Michael G. Rossmann*, Edward Arnold*, John W. Erickson*, Elizabeth A. Frankenger*, James P. Griffith*, Hans-Jürgen Hecht*, John E. Johnson*, Greg Kamer*, Ming Luo*, Anne G. Mosser†, Roland R. Rueckert‡, Barbara Sherry‡ & Gerrit Vriend*

* Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

† Biophysics Lab, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA

We report the first atomic resolution structure of an animal virus, human rhinovirus 14. It is strikingly similar to known icosahedral plant RNA viruses. Four neutralizing immunogenic regions have been identified. These, and corresponding antigenic sequences of polio and foot-and-mouth disease viruses, reside on external protrusions. A large cleft on each icosahedral face is probably the host cell receptor binding site.

PICORNAVIRUSES comprise one of the largest families of viral pathogens and cause serious diseases in humans and other animals, such as the common cold, poliomyelitis, foot-and-mouth disease and hepatitis. They are among the smallest RNA-containing animal viruses¹, with relative molecular mass (M_r) $\sim 8.5 \times 10^6$ and containing $\sim 30\%$ by weight RNA. Their external diameter is roughly 300 Å and they form icosahedral shells. Picornaviridae have been subdivided into four genera on the basis of their buoyant density, pH stability and sedimentation coefficients: enterovirus (for example, polio, hepatitis A and coxsackie viruses), cardiovirus (for example, encephalomyocarditis and Mengo viruses), aphthovirus (for example, foot-and-mouth disease virus) and rhinovirus. They differ also in the number of known serotypes. For instance, there are 3 known serotypes for polioviruses, 7 for foot-and-mouth disease viruses (FMDV) and at least 89 for human rhinoviruses (HRV). Accordingly, it has been possible to produce effective vaccines for poliomyelitis and, with greater difficulty, for foot-and-mouth disease, but not for the common cold.

Picornavirions contain 60 protomers, each composed of 4 structural proteins VP1, VP2, VP3 and VP4 (for nomenclature see ref. 2). The M_r s of these proteins in HRV14 are 32,000, 29,000, 26,000 and 7,000, respectively. The capsid protein VP0 is cleaved into its components VP4 and VP2 only in the final stages of assembly³. The cleavage occurs at an Asn-Ser peptide in HRV14, and hence is not affected by the viral protease whose specificity is for Gln-Gly peptides. The assembly of the capsid occurs in a series of steps culminating in the insertion of RNA into capsids to produce mature virions with the concomitant cleavage of VP0^{1,3}.

The virions contain a single, positive strand of RNA which is translated into a single polypeptide and then processed stepwise into its component proteins⁴. The gene order is essentially the same for HRV, FMDV, poliovirus and encephalomyocarditis virus (EMCV). The initial cleavage of capsid precursor protein from the polypeptide is probably caused by a host cell protease, but subsequent cleavages are mainly dependent on the release of viral protease excised from the polypeptide⁵.

Considerable effort has been devoted to mapping topological relationships among VP1, VP2, VP3 and VP4 within the capsid using chemical labelling of the surface of intact particles^{6,7}, treatment with crosslinking reagents⁸, reaction with specific antibodies and crosslinking with ultraviolet light⁹. The consensus

is that VP1 is the most external and immuno-dominant protein, whereas VP4 is inaccessible from the outside but can be cross-linked with RNA on the interior. Heat treatment or mild denaturing agents cause a conformational change in the capsid, thus altering the response to antisera. Surprisingly, the internal capsid protein VP4 can dissociate and escape from the capsid during antigenic conversion¹.

The RNA, and hence by inference the polypeptide, has been sequenced for all three strains of poliovirus, for various FMDV strains, for two strains of rhinovirus¹⁰⁻¹², for EMCV and for hepatitis A virus. Protein-protein comparisons between HRV, poliovirus, EMCV, FMDV and hepatitis A virus sequences show that HRV and poliovirus are closely related, whereas sequence homology of HRV to EMCV, FMDV or hepatitis A virus is not immediately obvious, particularly for the structural proteins.

X-ray diffraction studies of crystalline picornaviruses have produced only limited information. Coxsackievirus crystals and poliovirus crystals¹³ were reported a long time ago and rhinovirus strain 1A crystals a little later. However, the technical problems involved in a complete structure determination were not solved until the elucidation of the small plant RNA viruses tomato bushy stunt virus (TBSV)¹⁴, southern bean mosaic virus (SBMV)¹⁵ and satellite tobacco necrosis virus (STNV)¹⁶. This encouraged the renewal of the crystallographic study of poliovirus¹⁷ and stimulated work on rhinovirus¹⁸ as well as Mengo virus. It was shown¹⁸ that rhinovirus and poliovirus

Table 1 Data used in structure determination

Resolution range (Å)	No. of unique observations ($F^2 > 1\sigma(F^2)$) for each data set and percentage of possible total					
	Native		1 mM KAUCN ₂		5 mM KAUCN ₂	
	No.	%	No.	%	No.	%
∞-30	358	63	331	57	231	40
30-15	3,445	86	2,847	70	1,850	46
15-10	9,109	84	7,757	70	5,029	46
10-7.5	17,532	83	14,641	68	9,486	45
7.5-5.0	70,043	81	57,577	65	34,046	39
5.0-3.5	184,864	78	138,477	58		
3.5-3.0	133,101	63	83,852	39		
3.0-2.75	60,629	36	25,271	15		
2.75-2.6	26,834	11	8,175	3		
Total	509,915	58	338,928	39	50,641	41
No. of film packs	83		48		11	
R-factor (%)	11.0		12.9		12.0	

$$R = \frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_h} \times 100$$

where I_h is the mean of the I_{hi} observations of reflection h .

† Present addresses: Department of Physical Biochemistry, AP-9A D-47E, Abbott Laboratories, Abbott Park, North Chicago, Illinois 60064, USA (J.W.E.); Department of Agronomy, Purdue University, West Lafayette, Indiana 47907, USA (E.A.F.); F. G. Roentgenstrukturanalyse, Universitaet Wuerzburg, Zentralbau Chemie, Am Hubland, D-8700 Wuerzburg, FRG (H.-J.H.).

crystals can be roughly isomorphous, suggesting close similarities of the viral capsids, a result later supported by sequence homologies.

Structure determination

HRV14 crystals were prepared as described by Arnold *et al.*¹⁸. The crystals are cubic with the dimension $a = 445.1 \text{ \AA}$ and belonging to space group $P2_13$. There are four particles per crystal unit cell with each virion situated on a crystallographic 3-fold axis along the body diagonal. Thus, one-third of the virus (or 20 icosahedral asymmetrical units) is in the crystallographic asymmetrical unit. A rotation function¹⁹ gave the particle orientation¹⁸ and packing considerations showed that the particle had to be near (0, 0, 0) or (1/4, 1/4, 1/4) given the choice of origin as defined in the *International Tables for Crystallography*.

The data used for the results given here were collected at the Cornell High Energy Synchrotron Source (CHESS). A new crystal was used for every exposure. A total of 83, 0.3°-oscillation film packs were eventually included in the native data (Table 1). Surveys for suitable isomorphous heavy-atom derivatives eventually produced two related compounds, 1 mM $\text{KAu}(\text{CN})_2$ and 5 mM $\text{KAu}(\text{CN})_2$. Full high-resolution data were collected of the former, but only partial low resolution (0.6 to 0.8° oscillation angles) of the latter.

The mean differences between heavy-atom derivatives and the native data set were rather small in proportion to the size of the native amplitudes (8% for 1 mM $\text{KAu}(\text{CN})_2$ and 12% for 5 mM $\text{KAu}(\text{CN})_2$) but also the size of the differences increased at a resolution $> 7 \text{ \AA}$. Thus, there seemed to be little substitution and some concern about lack of isomorphism, particularly at higher resolution.

Systematic vector search procedures were applied to 6 Å difference Pattersons. The first three-dimensional searches were based on locating the self-Patterson vectors within an icosahedral virion given its known orientation. Plausible sites were then explored further using partial four-dimensional searches in which the virion position was slightly varied and cross-Patterson vectors between particles were also considered. The results clearly showed a single site, A, on the icosahedral 3-fold axes at a radius of 111.0 Å for the 1 mM $\text{KAu}(\text{CN})_2$ compound and a possible substitution at the same site in the 5 mM $\text{KAu}(\text{CN})_2$ derivative. The best particle position was found to be at $x = y = z = 0.0006$. Three-dimensional searches were then conducted for a second site assuming the particle position as well as the A site, yielding site B in the 5 mM but not in the 1 mM $\text{KAu}(\text{CN})_2$ derivative. The heavy-atom parameters for both heavy-atom data sets were used for determining phases to 6 Å resolution. The resultant electron-density map, averaged over the 20 different non-crystallographic asymmetrical units, showed clearly the viral protein envelope between ~100 and 150 Å radius.

The 6 Å map was improved with two cycles of molecular replacement averaging and back-transformation^{20,21}. A difference map of the two Au heavy-atom data sets with respect to the native data, based on the improved phases, showed an unambiguous third site, C, in the 5 mM $\text{KAu}(\text{CN})_2$ derivative. The heavy-atom parameters (Table 2) were used to compute phases from 25 to 5 Å resolution. The resultant map was the starting point for five cycles of molecular replacement. The electron-density map was exceptionally clear, showing that there are probably three β -barrels per icosahedral asymmetrical unit.

Phase extension beyond 5 Å resolution to 3 Å was performed in 20 steps. In each step, the resolution was extended by three reciprocal lattice points using phases obtained by back-transforming the previous averaged map. Two to three molecular replacement real-space averaging cycles were usually sufficient to improve the R -factor to $< 30\%$ and correlation coefficient to > 0.5 in the current outermost resolution shell (Fig. 1). Attempts at increasing the resolution in larger steps led to erroneous phases, probably the result of satisfying the non-crystallographic symmetry constraints in isolated outer regions of reciprocal space independently of the origin, enantiomorph or Babinet's

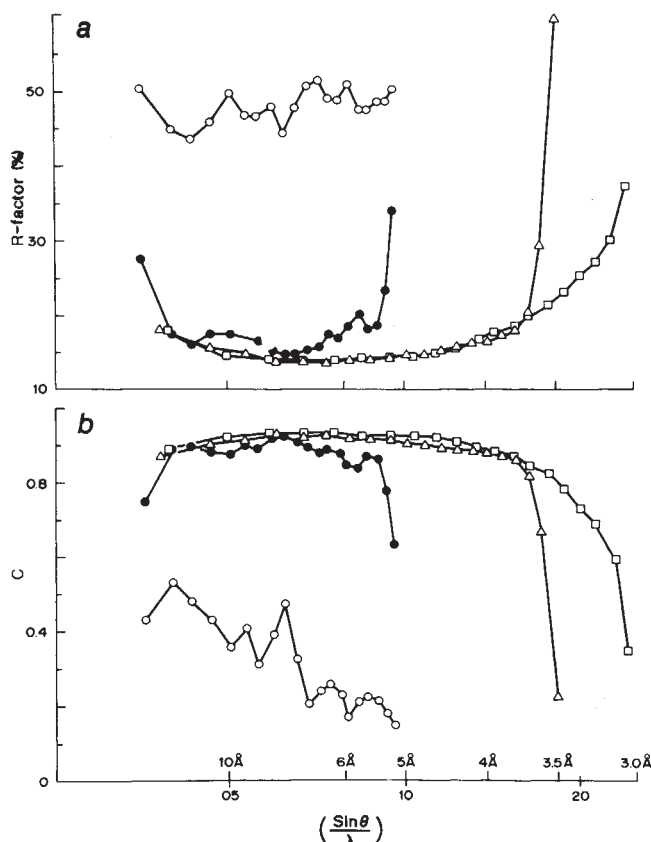


Fig. 1 Phase determination from 5 to 3 Å resolution by extension from 5 Å in small incremental steps of resolution. Shown are the correlation coefficients (C) and R -factors obtained from the initial 5 Å double isomorphous replacement determination ($\circ$), the 5-Å resolution refined phases, cycle 5 ($\bullet$), the 3.5-Å resolution extension, cycle 40 ($\triangle$) and the final 3.0-Å resolution results, cycle 60 ($\square$). Both R and C measure the agreement of the observed structure amplitudes and those calculated by Fourier back-transformation of the averaged cell. *a*, R -factor; *b*, C , where

$$C = \frac{\sum (\langle F_h \rangle - |F_0|) (\langle F_h \rangle - |F_c|)}{[\sum (\langle F_h \rangle - |F_0|)^2 (\langle F_h \rangle - |F_c|)^2]^{1/2}}$$

and

$$R = \frac{\sum (|F_0| - |F_c|)}{|F_0|}$$

Here $|F_0|$ and $|F_c|$ are observed and calculated structure amplitudes, respectively placed on the same relative scale for each local resolution shell; $\langle F_h \rangle$ is the mean observed amplitude in that shell.

inversion established at lower resolution. The particle envelope was defined by an external and internal sphere of 163 and 104 Å radius, respectively. Density outside the envelope was set to zero. The external sphere was truncated by planes, tangential to the map interparticle contacts, at a radial distance of 157 Å to avoid particle overlap, while allowing all the protein density to be included within the envelop. Electron densities used for averaging were computed on grids with spacing less than one-fifth of the resolution of the data.

When phases had been extended to 3.5 Å resolution, after 40 cycles of molecular replacement, a map of extremely good quality was calculated. Most residues could be easily recognized and the map interpreted fully. The final electron-density map was averaged and displayed in sections at 1-Å intervals perpendicular to an icosahedral 2-fold axis. A mini-map of the 3.5-Å resolution electron density was used for the original chain tracing and amino-acid identification. An atomic model was then built into the final 3.0-Å resolution map using an Evans and Sutherland PS300 computer graphics system and the FRODO program²². Most of the carbonyl oxygens and some water solvent molecules are readily recognizable.

Phase determination based on non-crystallographic symmetry was first suggested by Rossmann and Blow^{19,23}. It was used to compute very low-resolution maps of SBMV and polyoma virus, and for limited phase extension in the determination of STNV and haemocyanin. However, there was considerable concern that the molecular replacement method could fail in extending the phases all the way from low resolution (5 Å) to high resolution (3 Å). The high quality of the map reflects the precision of the diffraction data and the power and exactness of the non-crystallographic constraints as opposed to the approximations required in an isomorphous replacement phase determination. The success of these techniques holds promise for a rapid series of structural determinations of viruses and other biological assemblies of high symmetries, by reducing the dependence on finding satisfactory isomorphous heavy-atom derivatives.

Structure

The particle consists of an icosahedral protein shell (Fig. 2a) surrounding an RNA core. The lack of visible structure in the central cavity results from the random orientation of the asymmetrical RNA molecule. Both the tertiary fold of the VP1, VP2 and VP3 polypeptide chains and their quaternary organization within the HRV14 capsid are very similar to the two published high-resolution structures of $T=3$ (180 identical subunits per capsid) RNA plant viruses, TBSV¹⁴ and SBMV¹⁵. Although the subunit organization within the icosahedron is somewhat different, a similar tertiary structure has also been found for $T=1$ STNV (60 subunits per capsid)¹⁶. The radial position and orientation of structurally equivalent C_α atoms of HRV14 and SBMV generally agree to better than 3 Å relative to the icosahedral symmetry axes. In the plant viruses, the three quasi-equivalent subunits A, B and C have the same amino-acid sequence but cannot have identical geometrical environments²⁴. For SBMV, the first 63 of the 260 amino acids per subunit are associated with the RNA and therefore do not have icosahedral symmetry. In the C subunit, unlike the A and B subunits, residues 38–63 of SBMV are ordered with their ends forming a 'β-annulus' about the icosahedral 3-fold axes (Fig. 2b). Residues 64–260 of SBMV are referred to as the 'shell' domain and form an eight-stranded anti-parallel β-barrel (Fig. 3a). The β-barrels in HRV14, like those in SBMV, are wedge-shaped, with the thin end (left in Fig. 3) pointing towards the 5- or 3-fold (quasi-6-fold) axes. The alignment of the VP1, VP2 and VP3 chains of several picornaviruses is shown relative to the capsid protein of SBMV in Fig. 4.

The three larger capsid proteins VP1, VP2 and VP3 in HRV14 are oriented and situated at essentially the same radius and position as the A, C and B subunits in SBMV, respectively. The capsid proteins of HRV14 are related by a pseudo-3-fold axis, analogous to the quasi-3-fold axis in SBMV and TBSV, at the asterisk in Fig. 2b. However, VP1 and VP3 have additions at their amino- and carboxy-termini. The amino ends, as in plant viruses, are in contact with the internal RNA cavity, but unlike plant viruses, they are more acidic than basic and have only a very few amino-terminal residues 'disordered' (lacking icosahedral symmetry). The first 64 of the 73 amino-terminal residues of VP1 reside under VP3, whereas the first 42 of the 71 amino-terminal residues of VP3 are under VP1. Thus, the predominant positions of VP1 and VP3 at the RNA-protein interface are

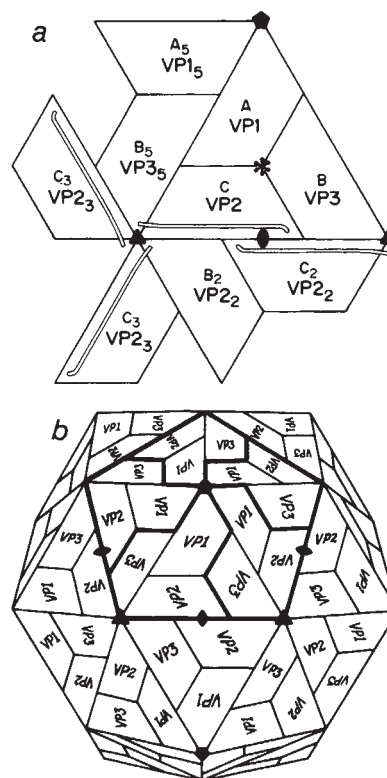


Fig. 2 Relation of the pseudo-equivalent VP1, VP2 and VP3 subunits to the quasi-equivalent subunits A, C and B in TBSV and SBMV. *a*, Icosahedral capsid; the thickly outlined VP1, VP3, VP2 unit corresponds to the 6S (VP1, VP3, VP0) protomer and the 15-mer cap to the 14S pentamer observed in assembly experiments. *b*, Icosahedral asymmetrical unit showing the ordered amino-terminal arm βA present only in the C subunit of the plant viruses and VP2 of HRV14. The amino-end of the arm interacts with two other VP2 arms across the 3-fold axis, whereas the carboxy-end of the arm interacts with a VP2 across the 2-fold axis. Asterisk indicates the position of the quasi-3-fold axis in SBMV and TBSV analogous to the pseudo-3-fold axis in HRV14. Subscripts designate the symmetry operation required to obtain the given subunit from the basic triangle.

exchanged relative to their positions at the exterior surface (Fig. 5).

The first 25 of the 69 residues of the internal structural protein VP4 are not seen in the electron-density map, implying that they lack icosahedral symmetry. VP4 is an extended polypeptide chain, positioned internally below, but in contact with, VP1 and VP2. Its visible amino end surrounds the 5-fold axis. The carboxy ends of VP1 and VP3 are external and function in part to associate proteins within a protomer (Figs 2a, 5).

Large sequence insertions relative to the typical shell domain (Fig. 4) form protrusions on VP1, VP2 and VP3 and create a deep cleft or 'canyon' on the viral surface. The canyon separates the major part of five VP1 subunits (in the 'north') clustered about a pentamer axis from the surrounding VP2 and VP3 subunits (Fig. 6) (in the 'south'), thus forming a moat around

Table 2 Heavy-atom parameters used for double isomorphous replacement phasing

Derivative	Site	Relative occupancy	No. of sites per virion	Fractional cell coordinates			Orthogonal coordinates (Å) with respect to icosahedron			Polar coordinates with respect to icosahedron			Chemical site of attachment	
				x	y	z	P	Q	R	R (Å)	(ψ°)	(φ°)		
1 mM KAu(CN) ₂	A	80.4	20	0.2389	-0.0186	-0.0739	0.0	39.8	104.2	111.5	69.1	-90.0	VP2	Cys 7
5 mM KAu(CN) ₂	A	64.2	20	0.2368	-0.0185	-0.0732	1.0	38.7	103.5	110.5	69.5	-89.5	VP2	Cys 7
	B	33.7	60	0.2637	0.0823	-0.0206	43.7	13.9	113.0	122.0	83.0	-69.0	VP1	Cys 69
	C	24.7	60	0.2606	0.0004	0.0389	6.1	-10.2	116.8	117.0	95.0	-87.0	VP2	Cys 248

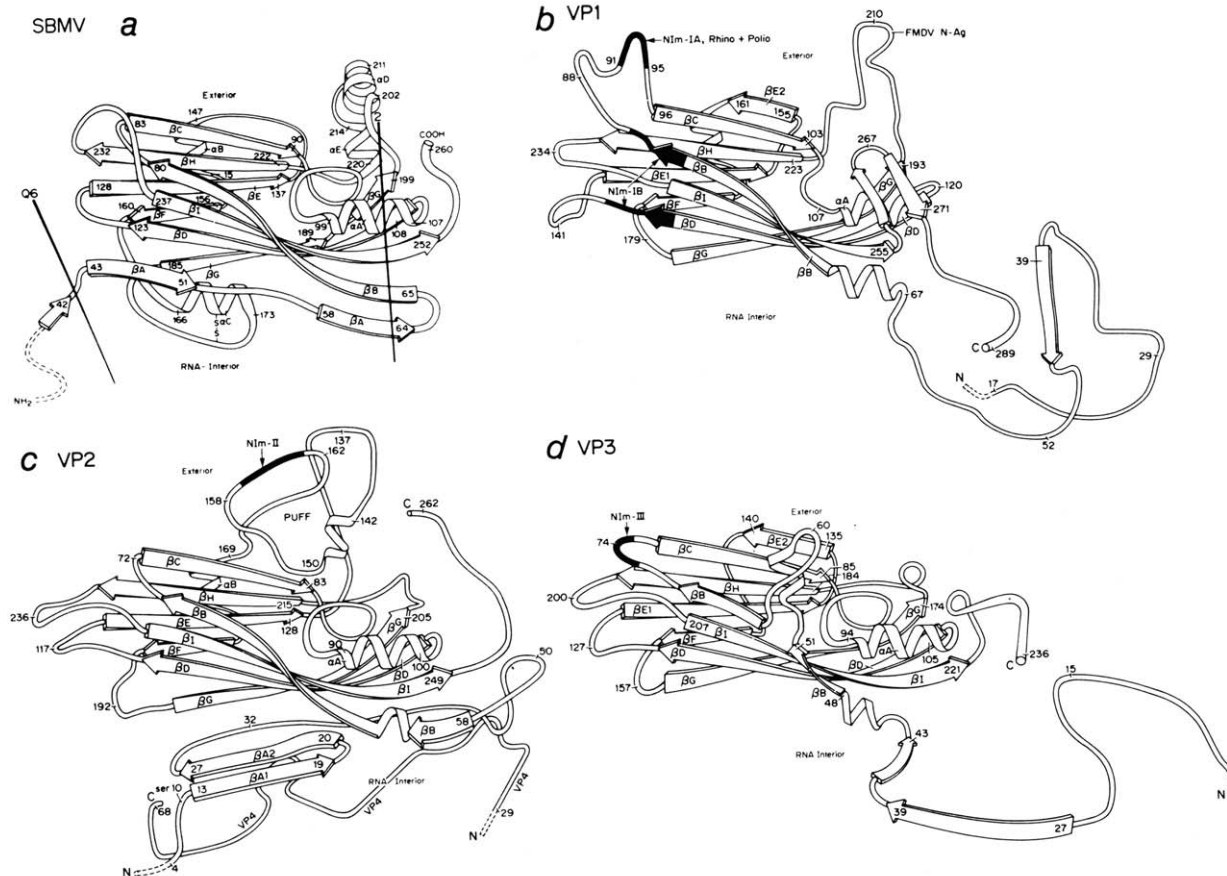


Fig. 3 Diagrammatic drawings showing the polypeptide fold of SBMV and of each of the three larger capsid proteins of HRV14. The nomenclature of the secondary structural elements is derived from that of SBMV⁴⁰. Amino-acid sequence numbers, appropriate for each protein, are also shown. *a*, SBMV; *b*, VP1 of HRV14; *c*, VP2 and VP4 of HRV14; *d*, VP3 of HRV14. There are four excursions of the polypeptide chains towards the wedge-shaped end. Each excursion makes a sharp bend or 'corner'. The most exterior (top left) corner is formed between the β -sheets βB and βC , the second corner down is formed between βH and βI , the third corner is between βD and βE and the most internal corner connects βF and βG . The βF - βG corner is the site of a 25-residue insertion in SBMV, including the α -helix αC , that is not present in any of the viral proteins of picornaviruses, nor in TBSV or STNV. (Adapted from a drawing of SBMV by Jane Richardson).

the VP1 protrusions on the 5-fold axis. The south canyon walls are lined with the carboxy-terminal ends of VP1 and a large sequence insertion in VP1 corresponding to helices αD and αE in the equivalent SBMV capsid protein. The north canyon wall is partially lined with the carboxy terminus of VP3. VP2 is hardly associated at all with the canyon, whereas VP1 is the major contributor to the residues lining the canyon. The canyon is 25 Å deep and 12–30 Å wide.

Because of the elaborations which VP1 has on the surface relative to VP2 and VP3, its overall shape is that of a kidney, with the depression forming a large part of the canyon (Fig 6*a, b*). The first 16 residues of VP1 (Fig. 3*b*) are not seen in the electron-density map. The shell domain of VP1 in HRV14 starts at residue 74. The small sequence insertion between βB and βC in rhinovirus and poliovirus is not found in FMDV. This loop forms a major immunogen in HRV14, NIm-IA and poliovirus. (The previously reported²⁵ N-Ag1 and N-Ag11 have been renamed NIm-IA and NIm-III to identify them as neutralizing immunogens and to indicate their major coat protein locations on VP1 and VP3, respectively.) Five carbonyl oxygens of 5-fold-related Asn 141 residues converge on the vertex to chelate a presumably cationic ligand on the 5-fold axis. The residues in VP1 of HRV14 that are analogous to αD and αE helices in SBMV protrude to the surface and form part of the south rim of the canyon, but do not form helices in HRV14. The most external portion of this segment contains the major antigenic site of FMDV and has been predicted to form an α -helix²⁶. The carboxy-terminal 23 residues of VP1 line the south rim of the canyon in HRV14. There is little conservation of these residues among picornaviruses.

Table 3 Amino acids associated with each of the four major immunogenic sites in HRV14

	Amino-acid number	Wild type	Observed mutations
NIm-IA			
VP1	91	D	→ A E G H N V Y
VP1	95	E	→ G K
NIm-IB			
VP1	83	Q	→ H
VP1	85	K	→ N
VP1	138	D	→ E G
VP1	139	S	→ P
NIm-II			
VP2	158	S	→ F
VP2	159	A	→ V
VP2	161	E	→ D V* D* K
VP2	162	V	→ M A*
VP1	210	E	→ D*
VP2	136	E	→ G
NIm-III			
VP3	72	N	→ I
VP3	75	R	→ G K M
VP3	78	E	→ K V
VP1	287	K	→ I
VP3	203	G	→ D

Substitutions in mutants selected for resistance to neutralization by monoclonal antibodies.

* A double mutation to V161 and A162 occurred in two cases; a double mutation to D161 and D210 occurred in a third case.

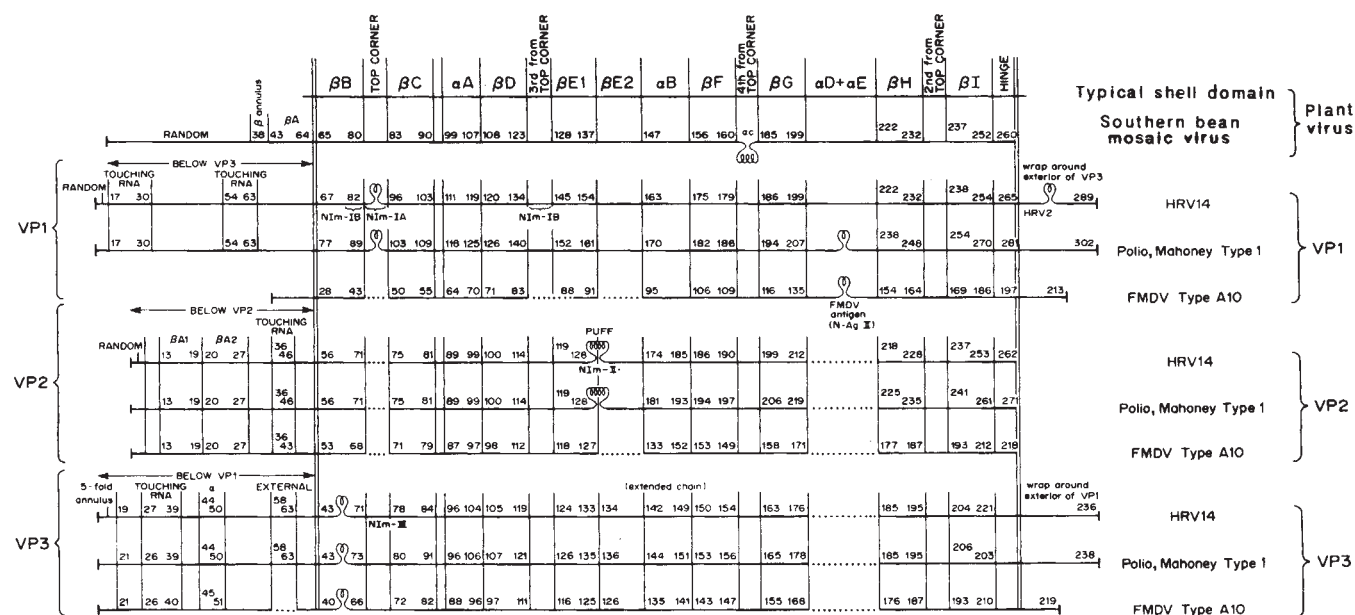


Fig. 4 Alignment of the three larger viral proteins VP1, VP2 and VP3 with a typical plant viral capsid protein as, for instance, the SBMV coat protein. The symbols $\odot$ and ... represent insertions and deletions with respect to the typical plant virus capsid protein. Immunogenic sites found in HRV14 are also shown. Three different picornaviruses are considered. The assignments of HRV14 are derived from the electron-density map and the known sequence of its RNA. The alignment of polio type 1 Mahoney was deduced from the amino-acid homology. The alignment of FMDV type A10, determined from amino-acid homology with all available picorna sequences, is uncertain from residues 1 to 190 in VP1.

The first three residues of VP2 (Fig. 3c) are not seen in the electron-density map. The proximity of Ser 10 in VP2 to the carboxy-terminus of VP4 suggests that the VP0 cleavage which occurs during the virus maturation step is autoproteolytic. There is no histidine next to Ser 10 and the carboxy end of VP4. However, nucleotide bases of the RNA could act as proton acceptors in the autocatalysis. Thus, the insertion of RNA into the growing capsids could trigger the cleavage of VP0 into VP2 and VP4. The α D and α E helices are absent in VP2 (Fig. 3c). There is a large 43-residue insertion in the VP2 position, corresponding to β E2 of VP1 and VP3 and forming an external mushroom-shaped 'puff'. This is positioned next to the VP1 elaborations, associated with the major antigenic site in FMDV, which line the south canyon wall (Fig. 6). The most external residues of this puff correspond to NIm-II of HRV14. In contrast to VP1 and VP3, the carboxy-terminus of VP2 has no extension beyond the shell domain.

All residues of VP3 (Fig. 3d) can be seen in the electron-density map. The 26 amino-terminal residues form a 5-fold β -barrel about the pentamer axis analogous to the β -annulus¹⁴ about trimer axes in SBMV and TBSV. This 5-fold annulus extends down into the RNA to a radius of 111 Å. The polypeptide emerges from the β -annulus and circles around the base of the VP1 shell domain, probably making extensive contact with the RNA in the central cavity. It then emerges on the viral surface near residue 61 and enters the shell domain at residue 72. The top corner of the VP3 shell domain, between β B and β C, is the NIm-III site of HRV14, structurally equivalent to NIm-IA in VP1. The external helices α D and α E of SBMV are absent, as in VP2. When the shell domains of VP3 and VP1 are superimposed, much of the amino- and carboxy-terminal arms are also structurally equivalent.

Immunogenic sites on HRV14

The immunological response to a virus is a major defence against disease in animals. Antibodies can bind to viruses but they do not necessarily neutralize infectivity. Despite the 60-fold equivalence of each potential binding site on the virus, as few as four neutralizing antibodies per virion can be sufficient to inhibit infectivity of poliovirus²⁷. Neutralizing antibodies usually change the isoelectric point of the picornaviruses^{28,29},

indicating that a conformation change frequently accompanies neutralization. Antibodies may neutralize by interfering with cell attachment, membrane penetration or virus uncoating. Antibodies that bind to poliovirus may require bivalent attachment for neutralization of the virus^{27,30}.

Amino-acid residues within the major neutralization immunogens of HRV14 have been identified by Sherry and Rueckert²⁵ and by Sherry *et al.*³¹. (The region represented by a synthetic peptide need not be immunogenic to a corresponding sequence on a viral surface; hence, by using peptide immunization, antibody specificities can be generated that cannot be obtained by other methods. Thus, an immunogen is here defined as a site on the intact virus that has been demonstrated to generate neutralizing antibodies.) They isolated mouse hybridoma lines which secrete monoclonal antibodies that neutralize HRV14 and each was then used to select several viral mutants resistant to neutralization by that antibody. Finally, every monoclonal antibody was assayed for its ability to neutralize the mutants. The results revealed four major immunogenic neutralization sites, each one composed of overlapping epitopes where a given mutant was resistant to many or all of the antibodies directed against that site. The amino-acid residues defining the four immunogens are summarized in Table 3. These results were obtained without knowledge of the three-dimensional structure. When, however, the electron-density map became available, it was immediately clear that the substitutions that could confer resistance to neutralization, regardless of their location in the amino-acid sequences, were localized into four distinct areas (NIm-IA, NIm-IB, NIm-II and NIm-III) corresponding exactly to the proposed immunogens. Moreover, these residues invariably faced outward towards the viral exterior (Figs 6a, 7). The large protrusion near NIm-III, consisting of residues 282–286 in VP1 and 58–63 in VP3, has not been shown to be antigenic for HRV14 but in FMDV residues associated with the carboxy-terminal end of VP1 have shown antigenicity (Table 4).

Relation to other antigenic sites

Table 4 summarizes results obtained in mapping the immunogenic and antigenic surfaces of other picornaviruses. There is a dominant immunogen in poliovirus corresponding to

Table 4 Evidence for antigenic regions in poliovirus and FMDV

Virus	Coat protein	Amino-acid no.	HRV14* amino-acid equivalent	Serotype	Ref.	Method	Position in HRV14 structure
Polio	VP1	11-17	deleted	1M	29	3	Within RNA
Polio	VP1	24-40	14-30	1M	46	5	Faces RNA
Polio	VP1	70-75	60-65	1M	29	3, 6	Faces RNA
Polio	VP1	61-80	51-70	1M	46	2	Faces RNA
Polio	VP1	86-103	77-97	1M	46	5	NIm-IB + NIm-IA
Polio	VP1	91-109	84-103	1M	46	2	NIm-IB + NIm-IA
Polio	VP1	100-109	93-103	1M	46	2	Part of NIm-IA
Polio	VP1	93-103	86-97	1M	29	2, 3, 6	NIm-IB(?) + NIm-IA†
Polio	VP1	93-104	86-98	1	47	4	NIm-IB(?) + NIm-IA†
Polio	VP1	93-100	86-93	3L	48, 49	1	NIm-IB(?) + part of NIm-IA†
Polio	VP1	141-147	135-141	1M,S	50	3	NIm-IB
Polio	VP1	161-181	154-174	1M	46	5	Partly exposed on canyon
Polio	VP1	182-201	175-193	1M	46	2	Buried on 4th corner down
Polio	VP1	222-241	212-225	1M	46	2	NIm-II
FMDV	VP1	141-160	210-228	O ₁ K	51	2	NIm-II
FMDV	VP1	144-159	211-227	O ₁ K	26	2	NIm-II
FMDV	VP1	145-168	211-236	A12	52	8	NIm-II
FMDV	VP1	146-154	211-223	O ₁ K	37	7	NIm-II
FMDV	VP1	169-179	237-247	A12	38, 52	8	NIm-IB central strand
Polio	VP1	270-287	254-272	1M	46	5	Partly exposed on canyon wall
FMDV	VP1	200-213	268-294	O ₁ K	51	2	NIm-III(?)
FMDV	VP1	200-213	268-294	O ₁ K	37	7	NIm-III(?)
Polio	VP2	162-173	161-170	1M	53	2, 3	NIm-II
Polio	VP3	71-82	70-81	1M	53	6	NIm-III

Serotype: 1M (type 1, Mahoney), S (Sabin), 3L (type 3, Leon), O₁K (type O₁, strain Kaufbeuren), A12 (type A, subtype 12). Numbers in methods column refer to the following techniques: (1) Monoclonal antibodies raised against intact virus select for resistant mutations; (2) synthetic peptides induce neutralizing antibodies; (3) synthetic peptides prime for high titre neutralizing response; (4) synthetic peptide competes with monoclonal antibody to inhibit neutralization; (5) synthetic peptides induce antibodies which bind virus but neutralize poorly; (6) neutralizing antisera or monoclonal antibodies bind peptide in an enzyme-linked immunosorbent assay; (7) deduced from ability or inability of protein fragments to induce neutralizing antibodies; (8) deduced from ability or inability of neutralizing monoclonal antibodies to bind protein fragments.

* Aligned by eye and by fitting to the shell domain structure.

† Uncertainty caused by somewhat arbitrary nature of computer alignments.

NIm-IA of rhinovirus. The absence of a corresponding immunogen in FMDV is explained by deletion of this loop from its VP1.

The dominant antigenic site in FMDV resides in a region homologous with NIm-II in HRV14 (Table 4). The VP1 contribution to this antigen has also been shown for HRV14 (Glu 210 of NIm-II in Table 3) and for poliovirus (residues 222-241, Table 4). In FMDV, however, the immunogenic puff on VP2 is absent and eight amino acids are inserted at the extreme surface of the α D- α E region of VP1. The resulting protrusion of NIm-II in FMDV would be unsupported by the puff. Therefore, this protrusion may occupy the space which in HRV14 is occupied by the puff. The inability of FMDV VP1 to elicit neutralizing antibodies³² is consistent with the lack of structural support for this NIm-II protrusion in the absence of the neighbouring VP2.

Poliovirus can also be neutralized by antibodies that bind to NIm-II or NIm-III (Table 4). Thus, there is an overall consistency in identifying at least three of the major HRV14 neutralization sites in poliovirus.

Many of the methods used to determine antibody-binding sites depend on the use of synthetic peptides as antigens (Table 4). Peptides associated with neutralizing antigenic regions do, in general, elicit antibodies that can neutralize the intact virus. In a significant number of cases (lines 1, 2, 3, 4, 8 and 13 of Table 4), however, the sequence in question lies far below the viral surface or is even buried in the RNA. This suggests that some peptides can elicit antibodies which subsequently bind to totally unrelated portions of the native virus. Alternatively, some of the results could be accounted for by conformational changes occurring during the isoelectric transition of the virus.

The canyon as receptor binding site

Despite the sequence and surface similarities of picornaviruses, they have different host and tissue specificity. Abraham and Colonna³³ have shown, using 24 rhinovirus serotypes in compe-

tion binding assays, that, although most serotypes recognize one receptor, a second smaller group recognizes a different receptor. HRV14 (sequenced by Callahan *et al.*¹¹ and by Stanway *et al.*¹⁰) belongs to the larger receptor group whereas HRV2 (sequenced by Skern *et al.*¹²) belongs to the smaller one. Krah and Crowell³⁴ have characterized some properties of HeLa cell receptors for group B coxsackieviruses. They found that concanavalin A (Con A) and other lectins adsorbed to receptors and inhibited virus attachment, a finding similar to that of Lonberg-Holm³⁵.

The presence of the 25-Å deep canyon, circulating around each of the 12 pentamer vertices (Fig. 6), suggests that this is the site for cell receptor binding. An antibody molecule, whose Fab fragment would have a diameter of the order of 35 Å, would have difficulty in reaching the canyon floor, its entrance being blocked by the canyon rim. Thus, the residues in the deeper recesses of the canyon would not be under immune selection and could remain constant, permitting the virus to retain its ability to seek out the same cell receptor.

Although retention of the canyon structure for all picornaviruses is to be expected, variation in the residues lining the canyon should be anticipated between viruses that attach themselves to different host cell receptors; that is, FMDV, polio, HRV14 and HRV2, all of which recognize different receptors, should exhibit some variation in the residues lining the canyon wall. Note that those parts of the carboxy-terminal ends of VP1 and VP3 which line the canyon walls are some of the least conserved amino acids among picornaviruses. As the topology of the canyon should be retained, the highly conserved, structurally equivalent sequences (MYVPPGAPNP starting at 151 of VP1 and AYTTPGARGP starting at 130 of VP3 for HRV14) in rhino, polio and FMD viruses situated near the floor of the canyon may be significant.

FMDV can be treated with trypsin, causing cleavage at residues between 138 and 154 of VP1. This causes the virus to

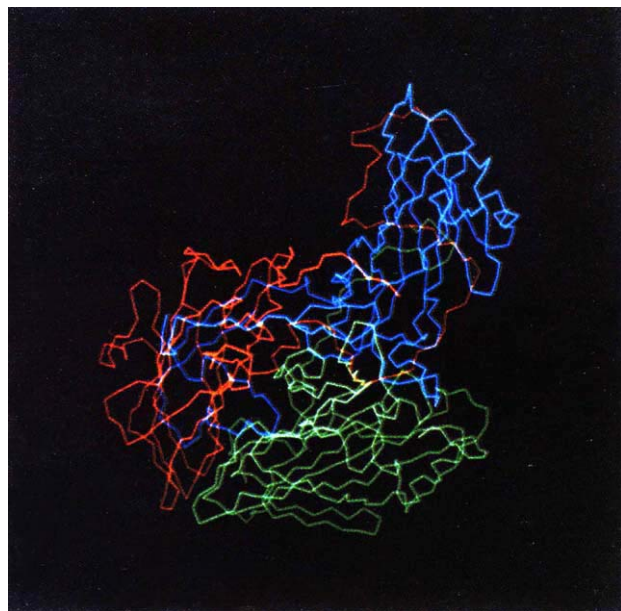


Fig. 5 Polypeptide tracing of VP1 (blue) and VP3 (red). The selected proteins are related by a pseudo-2-fold axis. Note the intertwining of the amino- and carboxy-terminal ends. Shown also are VP2 and VP4 in green. The chosen subunits are those most closely linked and probably correspond to the 6S protomer.

lose its ability to attach to cells and to stimulate neutralizing antibody³⁶⁻³⁸. The enzymatic cleavages occur in NIm-II on the α D- α E protrusion of VP1, a large loop which also forms part of the presumed host receptor binding site.

Argos *et al.*³⁹ have shown that there is a similarity in the folding topology of Con A and the shell domain of TBSV (and hence also to VP1), thus providing a possible rationalization of the competition experiments at the molecular level^{34,35}. The functional sugar binding site of Con A corresponds roughly to the NIm-IB site in VP1. Comparison of the amino-acid sequence between type 1 Mahoney (virulent) and type 1 Sabin (attenuated) shows that the differences are concentrated near the NIm-IB site. This raises the possibility that virulent and attenuated strains may have altered receptor specificity for carbohydrate recognition. Similarly, the deletion of HRV2 relative to HRV14 in β B near NIm-IB may, in part, be responsible for the different receptor specificity³³ of these two rhinoviruses.

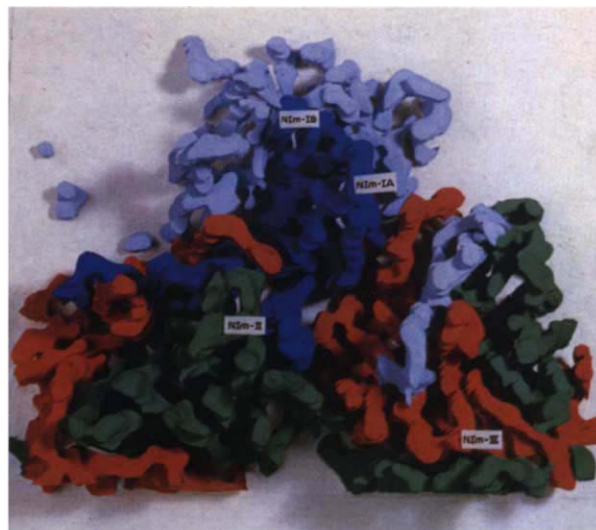
Neutralization

Because an antibody itself has a 2-fold axis, bivalent attachment could occur across icosahedral 2-fold axes. The distances between the nearest 2-fold-related immunogenic sites IA, IB, II and III are 120, 120, 50 and 60 Å, respectively. All other symmetrical counterparts of these antigens related by 2-fold axes are at least 170 Å apart. Lower and upper limits of the distance of the two antibody-binding sites on an immunoglobulin molecule are not well known but probably lie in the range of 50 to 180 Å²⁷. Hence, symmetrical bivalent attachment of antibodies to the known immunogens may be possible, although NIm-II and III are close to the probable limit. As Fab arms of antibodies are capable of some rotation, it is conceivable that bivalent asymmetrical attachment could also occur across 3- or 5-fold axes. The resultant torque may produce conformational changes sufficient to interfere with the normal process of infection.

Assembly

Assembly of picornaviruses¹ proceeds from 6S protomers of VP1, VP3 and VP0, via 14S pentamers of five 6S protomers, to mature virions. The final step involves inclusion of the RNA into empty capsids or partially assembled shells with simul-

a



b



Fig. 6 *a*, Balsa wood model of the 5 Å electron-density map showing the external portion of the virus. The balsa wood sheets are equivalent to a thickness of 1.5 Å and are perpendicular to an icosahedral 2-fold axis. The proteins VP1, VP2 and VP3 are coloured blue, green and red, respectively. The central VP1 protein is dark blue whereas the neighbouring, 5-fold-related, VP1 proteins are light blue. The positions of the major HRV14 immunogens are shown. Note the large canyon which is the proposed binding site of the host cell receptor. The base of the canyon is inaccessible to antibody binding, whereas the canyon rim is lined with antigenic sites. *b*, C α polypeptide chain tracing of VP1, VP2 and VP3 in two neighbouring protomers. Amino-terminal domains of VP1 and VP3 are omitted. Colour code is as in Fig. 5 and *a*.

taneous cleavage of VP0 into VP2 and VP4. Conversely, *in vitro* disassembly, produced by mild denaturation, proceeds via the expulsion of VP4 followed by the RNA¹.

Both the amino and carboxy ends of VP1 and VP3, are intertwined with each other (Fig. 5). Furthermore, if VP4 and VP2 are considered as VP0, then VP0 is also intertwined with VP1 and VP3, which strongly suggests that the 6S protomer is as shown in Fig. 2*a*. These protomers are themselves intertwined because of the 5-fold β -annulus formed by the amino ends of the VP3s and the proximity of the observed amino ends of VP4s to the 5-fold axis. Thus, the 14S pentamers closely correlate

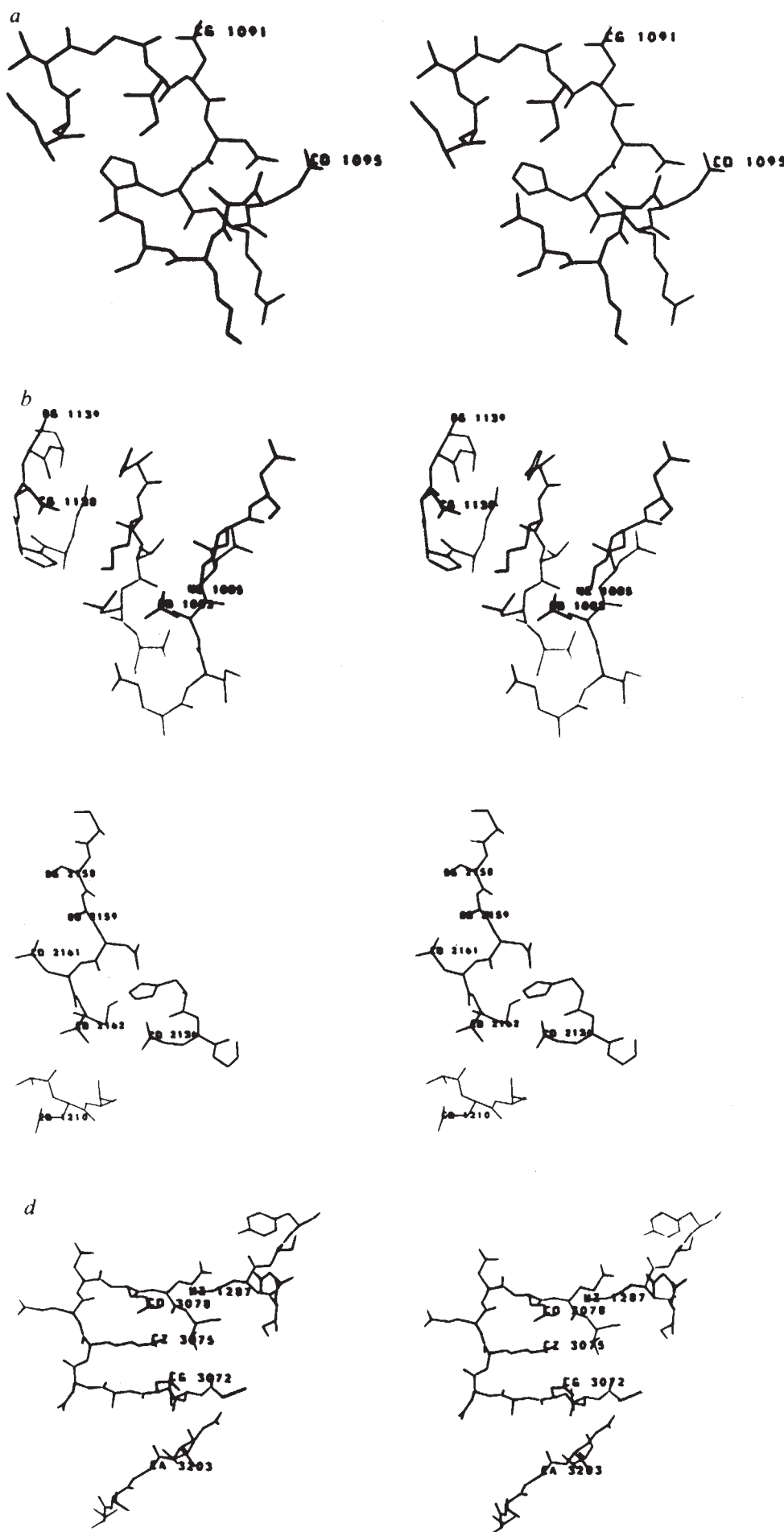


Fig. 7 Stereo view of each of the four antigenic sites on the HRV14 surface. The models were built with respect to the 3.0-Å resolution electron-density map on an Evans and Sutherland PS300 system using the FRODO program package. 1,000 has been added to residue numbers in VP1, 2,000 for residues in VP2 and 3,000 for residues in VP3. Identified atoms are C_α (CA), C_β (CB), C_γ (CG), O_γ (OG), C_δ (CD), C_ε (CZ) and N_ε (NZ). *a*, NIm-IA is an insertion between βB and βC at the top-most corner of the VP1 wedge. It is the most external portion of the complete virion, being 160 Å from the centre. Residues 91 and 95 are on the extreme external portion of the loop. *b*, NIm-IB on VP1 is at the carboxy ends of βB and βD, which are situated on either side of the amino-end of strand βI. No mutations conferring resistance have yet been found on the βI strand. The NIm-IB site is close to the 5-fold axis, a little below the canyon rim. *c*, NIm-II on VP2 is at the extreme outside of the puff, 155 Å from the viral centre. This immunogenic puff is next to the external loop formed by the sequence insertion in VP1 corresponding to αD and αE of SBMV. One of the residues (Glu 210 on VP1) that is associated with NIm-II is on VP1 (Table 3). *d*, NIm-III on VP3 is 149 Å from the viral centre to some extent in the shadow of a larger protrusion of VP3 (residues 58–63) and the carboxy end of VP1 (282–286). Residue Lys 287 on VP1, associated with this site (Table 3), points directly towards and is next to the other residues on VP3 associated with NIm-III.

with the observed structure, shown diagrammatically in Fig. 2a. Such an assembly sequence matches that observed in plant viruses, in particular that of SBMV, where the building blocks are dimers corresponding to VP1 and VP3₅, and where the formation of intermediates with 5-fold symmetry is considered to be a critical stage in the formation of $T=1$ and $T=3$ capsids^{40,41}.

The protein VP2, once cleaved from VP4, is globular and does not contact the other proteins extensively, although there are extensive solvent-accessible regions around VP2 continuing into VP1. This, as well as the extraordinarily internal heavy-atom sites on VP2 (Table 2), is consistent with loose binding of VP2 to the capsid. (The ability of the heavy atoms to penetrate deeply into the shell suggests cautious interpretation of data based on the use of small chemical labels for mapping viral surfaces). Disruption of pentamer-pentamer contacts, mediated by a slight reorientation of VP2 or its complete removal, could provide a port by which the VP4 and RNA can exit. Binding of a cell receptor in the canyon next to VP3 could facilitate this process, possibly accompanied by an isoelectric change.

The amino-ends of the capsid proteins are invariably associated with the RNA in both plant and animal viruses. However, in TBSV, SBMV and STNV they are basic, whereas in HRV14 they are mildly acidic. These properties may be significant for the initial events of assembly.

Evolution

Conservation of three-dimensional structure is almost invariably greater than conservation of amino-acid sequence homology⁴². Thus, structural comparison can be used to trace divergent evolution over longer time spans than is possible by amino-acid comparisons. Many structural comparisons have now been made and their probability for divergence assessed. These provide

benchmarks to which other comparisons can be related. Thus, the considerable similarity of VP1, VP2 and VP3 structures to those of the plant viruses leaves little doubt as to their divergence from a common ancestor. Possibly the corresponding gene for a protein such as Con A provided a useful tool for the formation of biological assemblies that could attach themselves to animal cell receptors. Franssen *et al.*⁴³ and Argos *et al.*⁴⁴ have also shown that the virally coded polymerase, the genomic protein VPg and the protease of polioviruses are homologous in sequence and gene order with cowpea mosaic virus. These comparisons have been extended to other viruses, for example, alfalfa mosaic virus, brome mosaic virus, cucumber mosaic virus, tobacco mosaic virus and Sindbis virus⁴⁵.

We apologise for the omission of many references to significant and relevant papers because of space limitations. We thank Sharon Wilder, Kathy Shuster, Bill Boyle, Jun Tsao, Gale Rhodes, Diana Delatore and Tim Schmidt for help in data collection and presentation; Keith Moffat, Wilfried Schildkamp, Robert Hunt, Don Bilderbeck, Aggie Sirrine and Boris Batterman at CHESS; Hans Bartunik and Klaus Bartels at DESY; Saul Rosen, John Steele, Tom Putnam and Paul Townsend at the Purdue University Computer Center; Richard Colonna, Ann Palmenberg, Jeffrey Bolin, Abelardo Silva, Ignacio Fita, Celerino Abad-Zapatero, R. Usha, M. V. Hosur, Cynthia Stauffacher, Patrick Argos and J. K. Mohana Rao for helpful discussions. The work was supported by an NIH grant, NSF grants for supercomputer time and computer graphics, a Purdue University Showalter Foundation grant and a contribution by Merck Sharp & Dohme Co. to M.G.R. and an ACS grant to R.R.R. CHESS is supported by an NSF grant and the macromolecular diffraction facility by an NIH grant. A postdoctoral Walter Winchell-Damon Runyon fellowship supported E.A. and a predoctoral NIH training grant supported B.S.

Received 3 June; accepted 30 July 1985.

- Rueckert, R. R. *Comprehensive Virology* Vol. 6 (eds Fraenkel-Conrat, H. & Wagner, R. R.) 131-213 (Plenum, New York, 1976).
- Rueckert, R. R. & Wimmer, E. *J. Virol.* **50**, 957-959 (1984).
- Jacobson, M. F., Asso, J. & Baltimore, D. *J. molec. Biol.* **49**, 657-669 (1970).
- Pallansch, M. A. *et al. J. Virol.* **49**, 873-880 (1984).
- Hanecak, R., Semler, B. L., Anderson, C. W. & Wimmer, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3973-3977 (1982).
- Lonberg-Holm, K. & Butterworth, B. E. *Virology*, **71**, 207-216 (1976).
- Beneke, T. W., Habermehl, K. O., Diefenthal, W. & Buchholz, M. *J. gen. Virol.* **34**, 387-390 (1977).
- Hordern, J. S., Leonard, J. D. & Scraba, D. G. *Virology*, **97**, 131-140 (1979).
- Wetz, K. & Habermehl, K. O. *J. gen. Virol.* **59**, 387-401 (1982).
- Stanway, G., Hughes, P. J., Mountford, R. C., Minor, P. D. & Almond, J. W. *Nucleic Acids Res.* **12**, 7859-7875 (1984).
- Callahan, P. L., Mizutani, S. & Colonna, R. J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 732-736 (1985).
- Skern, T. *et al. Nucleic Acids Res.* **13**, 2111-2126 (1985).
- Finch, J. T. & Klug, A. *Nature* **183**, 1709-1714 (1959).
- Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276**, 368-373 (1978).
- Abad-Zapatero, C. *et al. Nature* **286**, 33-39 (1980).
- Liljas, L. *et al. J. molec. Biol.* **159**, 93-108 (1982).
- Hogle, J. M. *J. molec. Biol.* **160**, 663-668 (1982).
- Arnold, E. *et al. J. molec. Biol.* **177**, 417-430 (1984).
- Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **15**, 24-31 (1962).
- Argos, P., Ford, G. C. & Rossmann, M. G. *Acta crystallogr.* **A31**, 499-506 (1975).
- Bricogne, G. *Acta crystallogr.* **A32**, 832-847 (1976).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **16**, 39-45 (1963).
- Caspar, D. L. D. & Klug, A. *Cold Spring Harb. Symp. quant. Biol.* **27**, 1-24 (1962).
- Sherry, B. & Rueckert, R. *J. Virol.* **53**, 137-143 (1985).
- Pfaff, E., Mussgay, M., Böhm, H. O., Schulz, G. E. & Schaller, H. *EMBO J.* **1**, 869-874 (1982).

- Icenogle, J. *et al. Virology* **127**, 412-425 (1983).
- Mandel, B. *Virology* **69**, 500-510 (1976).
- Emini, E. A., Jameson, B. A. & Wimmer, E. *Nature* **304**, 699-703 (1983).
- Emini, E. A., Ostapchuk, P. & Wimmer, E. *J. Virol.* **48**, 547-550 (1983).
- Sherry, B., Mosser, A. G., Colonna, R. J. & Rueckert, R. R. *J. Virol.* (in the press).
- Meloen, R. H., Briare, J., Wootmeyer, R. J. & Van Zaane, D. *J. gen. Virol.* **64**, 1193-1198 (1983).
- Abraham, G. & Colonna, R. J. *J. Virol.* **51**, 340-345 (1984).
- Krah, D. L. & Crowell, R. L. *J. Virol.* **53**, 867-870 (1985).
- Lonberg-Holm, K. *J. gen. Virol.* **28**, 313-327 (1975).
- Cavanagh, D., Sangar, D. V., Rowlands, D. J. & Brown, F. J. *J. gen. Virol.* **35**, 149-158 (1977).
- Strohmaier, K., Franze, R. & Adam, K. H. *J. gen. Virol.* **59**, 295-306 (1982).
- Baxt, B., Morgan, D. O., Robertson, B. H. & Timpone, C. A. *J. Virol.* **51**, 298-305 (1984).
- Sherry, B., Tsukihara, T. & Rossmann, M. G. *J. molec. Biol.* **15**, 169-179 (1980).
- Rossmann, M. G., Abad-Zapatero, C., Hermodson, M. A. & Erickson, J. W. *J. molec. Biol.* **166**, 37-83 (1983).
- Erickson, J. W., Silva, A. M., Murthy, M. R. N., Fita, I. & Rossmann, M. G. *Science* **229**, 625-629 (1985).
- Rossmann, M. G. & Argos, P. A. *Rev. Biochem.* **60**, 497-532 (1981).
- Franssen, H., Leunissen, J., Goldbach, R., Lomonosoff, G. & Zimmern, D. *EMBO J.* **3**, 855-861 (1984).
- Argos, P., Kamer, G., Nicklin, M. J. H. & Wimmer, E. *Nucleic Acids Res.* **12**, 7251-7267 (1984).
- Haseloff, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4358-4362 (1984).
- Chow, M., Yabrov, R., Bittle, J., Hogle, J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 910-914 (1985).
- Wychowski, C. *et al. EMBO J.* **2**, 201-2024 (1983).
- Evans, D. M. A., Minor, P. D., Schild, G. S. & Almond, J. W. *Nature* **304**, 459-462 (1983).
- Minor, P. D. *et al. Nature* **301**, 674-679 (1983).
- van der Werf, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 5080-5084 (1983).
- Bittle, J. L. *et al. Nature* **298**, 30-33 (1982).
- Robertson, B. H., Morgan, D. O. & Moore, D. M. *Virus Res.* **1**, 489-500 (1984).
- Emini, E. A., Jameson, B. A. & Wimmer, E. in *Modern Approaches to Vaccines* (eds Chanock, R. M. & Lerner, R. A.) 65-75 (Cold Spring Harbor Laboratory, New York, 1984).

A fast pulsar candidate in the globular cluster M28

M. J. Mahoney & W. C. Erickson

Clark Lake Radio Observatory, University of Maryland,
College Park, Maryland 20742, USA

Hamilton, Helfand and Becker¹ (hereafter termed HHB) have recently used the Very Large Array (VLA) telescope at 1,465 MHz to observe radio sources in globular clusters. These observations were primarily motivated by the suggestion² that millisecond pulsar progenitors could be identified with low-mass X-ray binaries (LMXBs), and the fact that the frequency of occurrence of LMXBs per unit mass of stars is ~ 100 times greater in the cores of globular clusters than in the field. HHB found a source within the core of M28, which we have observed at 30.9 and 57.5 MHz using the Clark Lake TPT synthesis telescope³. Our observations show that this source has a spectral index of -2.44 . Only pulsars have well-documented spectra which are as steep as this.

It has long been known that the number of radio sources near the positions of globular clusters is consistent with the number expected based on extragalactic source counts⁴. In their survey of globular clusters, HHB detected 21 radio sources within 10 core radii of the centres of the 12 clusters that they observed; only one cluster, M28, contained a source within 1 core radius. Its 1,465-MHz flux density was 1.4 mJy, which, when combined with an additional VLA observation that yielded an upper limit of 0.38 mJy at 5,000 MHz, implied a spectral index of less than -1.1 .

Although the number of sources detected in their survey is compatible with $\log(N)/\log(S)$ (log number/log flux density) counts, HHB estimated, based on millisecond pulsar beaming factors, intrinsic luminosities, and lifetimes, that approximately one fast pulsar would have been expected to appear in their survey. Thus the steep spectrum source in the core of M28, which we designate 1821-248, is potentially very interesting.

The observations reported here were obtained during the period 10-15 April 1985 using the Clark Lake Radio Observatory's TPT synthesis telescope. As it operates in the 15-125 MHz frequency range, the TPT is a unique instrument for the detection of continuum flux from steep spectrum sources such as millisecond pulsars⁵.

The results of observations from three nights are shown in Fig. 1. Because low-frequency observations are subject to ionospheric refraction, several sources from the preliminary Texas 365-MHz survey (J. N. Douglas, personal communication) are indicated on the maps to verify the position of M28. All of the sources in the 30.9-MHz field appear in the preliminary Texas survey except 1821-248 which, because of its steep spectrum, would be below the Texas flux density limit, and G7.7-3.7 which would be resolved by the Texas interferometer.

Figure 2 shows the spectrum of 1821-248 between 30.9 and 1,465 MHz; also shown are the upper limits found at 365 MHz from the Texas survey and the VLA upper limit at 5,000 MHz. The flux densities measured at Clark Lake are 16.9 ± 5.6 Jy at 30.9 MHz and 4.1 ± 1.3 Jy at 57.5 MHz; they have been referenced to the flux density scale of Kuhr *et al.*⁶. A least-squares fit to the spectrum of this source has a spectral index of -2.44 . This is essentially the same spectral index as that of the 6 ms pulsar 1953+290, which is believed to be a LMXB (ref. 7). Note that the Crab Nebula pulsar and the 1-ms second pulsar, 1937+214, each have similar spectral indices of -2.3 to -2.5 . If, in fact, 1821-248 is found to be a millisecond pulsar, it would strongly support the accretion spin-up model of binary systems for producing fast pulsars. Another implication is that millisecond pulsars should generally be easier to detect in globular clusters than along the galactic plane because of smaller dispersion measures.

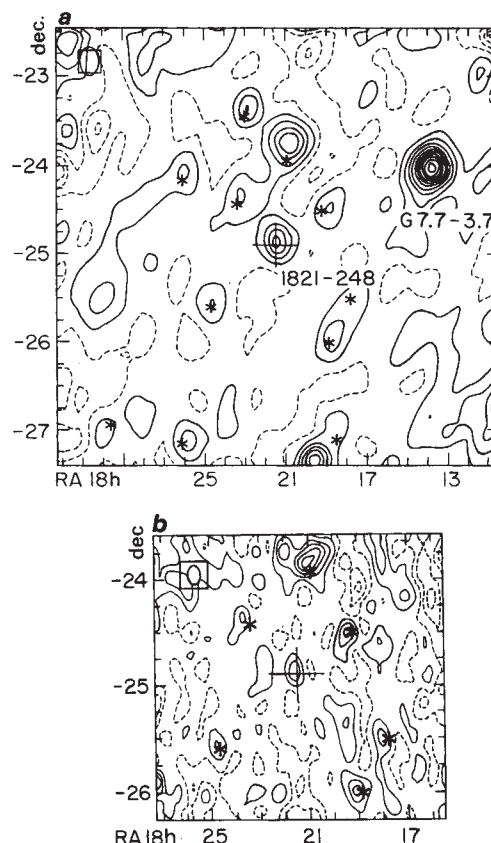


Fig. 1 *a*, A 30.9-MHz map of the M28 field. The map centre is right ascension (RA) (1950) = 18 h 21 min 30 s, declination (dec.) (1950) = $-24^{\circ} 54' 00''$. The beam size is 13.0×18.5 arc min and is shown in the box. Contour levels are 5% 15% up to 95% of the map maximum. The locations of several sources from the Texas 365-MHz survey are indicated for positional reference. The strong source to the north-west is the galactic supernova remnant G7.7-3.7. *b*, A 57.5-MHz map of the M28 field. The field-of-view of the TPT increases linearly with wavelength, so this map has been scaled to the same angular size as *a*. The beam size is 7.5×11.5 arc min and the contour levels are 10% 30% 50% 70% 90% of the map maximum.

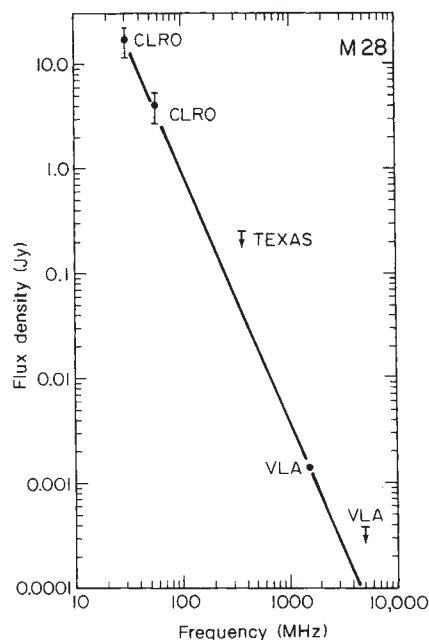


Fig. 2 The flux density spectrum of M28 derived from our observations and the VLA observations of HHB. The fitted line represents a spectrum of index -2.44 .

The present observations confirm that 1821–248, the core source in the globular cluster M28, has a very steep spectrum. This makes it a prime candidate for being a fast pulsar. Because of their relatively low intrinsic luminosity and their very steep spectra, pulsars are often difficult to detect as continuum sources at decimetre wavelengths; they are relatively easy to detect as continuum sources at decametric wavelengths with an instrument such as the Clark Lake TPT. Knowledge of the unusual spectrum of 1821–248 should encourage searches for pulsations and interplanetary scintillations from this object.

This work was supported by the NSF under grant AST 82-15463.

Received 14 May; accepted 20 June 1985.

1. Hamilton, T. T., Helfand, D. J. & Becker, R. H. *Astr. J.* **90**, 606–608 (1985).
2. Alpar, A., Cheng, A., Shaham, J. & Ruderman, M. *Nature* **300**, 728–730 (1982).
3. Erickson, W. C., Mahoney, M. J. & Erb, K. *Astrophys. J. Suppl.* **50**, 403 (1982).
4. Birkinshaw, M. & Downes, A. J. B. *Astrophys. J.* **258**, 154–160 (1982).
5. Erickson, W. C. *Astrophys. J. Lett.* **264**, L13–17 (1983).
6. Kuhr, H., Witzel, A., Pauliny-Toth, I. I. K. & Nauber, U. *Astr. Astrophys. Suppl.* **45**, 367–430 (1981).
7. Boriakoff, V., Bucher, R., Fauci, F., Turner, K. & Davis, M. *Millisecond Pulsars—Proc. HRAO Workshop No. 8*, 24–29 (1984).

Extra-low-frequency radiation from the polar electrojet antenna

R. Barr*, M. T. Rietveld†, H. Kopka†, P. Stubbe† & E. Nielsen†

* Geophysical Observatory, P.O. Box 2111, Christchurch, New Zealand

† Max-Planck-Institut für Aeronomie, 3411 Katlenburg-Lindau, FRG

There have been many recent reports on the detection of extra-low-frequency (ELF) radio waves in the immediate vicinity of powerful ionospheric high-frequency (HF) heating transmitters which are amplitude modulated at the ELF frequency^{1–4}. Results have also been presented which indicate that long path propagation (>3,500 km) of ELF signals from such heating facilities may be possible^{5,6}. However, no calibrations were provided with these measurements and it has thus not been possible to estimate the relative efficiency of these 'wireless' ELF generators compared with the more conventional long-wire antennas, although several theoretical estimates have been provided^{7–9}. Results are presented here of the simultaneous measurement at Lycksele and Kiruna, northern Scandinavia, of the ELF signals radiated by the polar electrojet antenna located near Tromsø. From these data, it has been possible to confirm the successful excitation of the Earth-ionosphere waveguide by the polar electrojet source and to make a first estimate of its radiated power—this is typically 1 W in the frequency range 1–1.5 kHz for HF effective radiated powers of 270 MW.

Figure 1 shows the region of northern Scandinavia pertinent to this study. The heating facility is located at Ramfjord, 15 km south of Tromsø, and the ELF receivers used for the measurements were located at Kiruna and Lycksele, 205 and 554 km respectively from Ramfjord. For all the data presented here, the heater was operated at a HF of 2.759 MHz in the x-mode with a maximum effective radiated power of 270 MW, with the beam directed vertically and the HF transmitter sine wave modulated at a frequency of 1.04 or 1.57 kHz. From 1730 (UT) on the 25 September 1982 to 1211 UT on the 26 September 1982, the transmitter was operated for 1 min in every 20, with 1.57 kHz modulation. From 1408 UT on the 26 September 1982 to 1529 UT on the 27 September 1982, the modulation frequency was 1.04 kHz. Gaps in the data represent operator rest periods rather than failure to receive a signal above the noise level.

The receiving system used synchronous detectors to measure the in-phase and quadrature components of the signal averaged over the 1-min transmission period. At Kiruna, the signals were

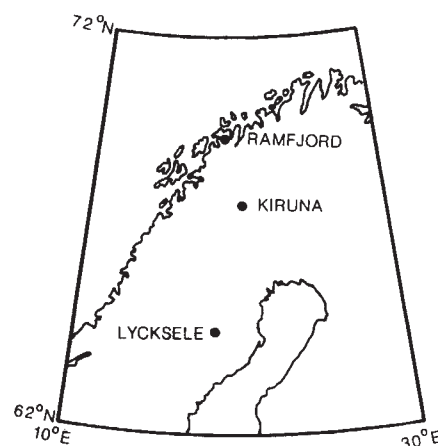


Fig. 1 Map of northern Scandinavia showing the location of the HF heating transmitter at Ramfjord and the ELF receiving sites at Kiruna and Lycksele.

received on a 2-m vertical rod antenna. At Lycksele, the signals were received on both a vertical rod and a loop antenna. The local ELF signals were also measured at Lavangsdalen, 17 km from Ramfjord, using loop antennas.

Figure 2 shows the signals received at 1.04 kHz (a–d) and 1.57 kHz (e–g) at Lavangsdalen, Kiruna and Lycksele. The signals recorded at Kiruna are highly variable, ranging in strength from 0.3 to 30 $\mu\text{V m}^{-1}$. As Kiruna is only 205 km from the signal source, it seems reasonable to assume that most of this variability is due to source effects. As might be expected from Earth-ionosphere-waveguide propagation¹⁰, the signals recorded at Lycksele, 554 km from the source, were much weaker, peaking at 10 $\mu\text{V m}^{-1}$. However, the amplitude variations at Kiruna and Lycksele seem to be correlated and an analysis shows the correlation coefficients to be 0.87 and 0.80 at 1.04 and 1.57 kHz respectively. This high correlation between two completely independent, spatially separated, recording systems shows that the ELF signals from the heating facility are propagating in the Earth-ionosphere waveguide and that the dominant mechanism for signal variability is provided by the source rather than the propagation medium. That the signal variability is provided by the source is also evident in similar amplitude fluctuations in the local ELF signals from Lavangsdalen, underneath the heated region. The Lavangsdalen data may also contain a slight diurnal variation due to the waveguide resonance effect³, especially at 1.57 kHz where a nocturnal enhancement may occur due to the decrease in the resonance frequency.

As two recording sites were available at differing distances from the transmitter, it was possible, using waveguide theory^{11,12}, to determine not only the effective power radiated by the polar electrojet source but also the propagation constant of the Earth-ionosphere waveguide. The effective power radiated by the polar electrojet source on 1.04 and 1.57 kHz during the 2-day period is shown in Fig. 2d, h. The great variability of the ELF power radiated by the polar electrojet antenna system is evident in Fig. 2, with ELF output powers ranging from <100 μW to >2 W. The upper value compares favourably with theoretical values of a few watts computed by Barr and Stubbe⁸ for a constant ionospheric electric field of 25 mV m⁻¹.

An attempt was made to correlate the logarithm of the ELF power radiated by the polar electrojet antenna with the 3-hourly K index recorded at Kiruna over the 2-day period of the observations. A correlation coefficient of 0.66 was obtained, supporting the work of Kapustin *et al.*¹ which suggested that ELF power radiated is related to magnetic activity. An attempt was also made to correlate the ELF electric field recorded at Kiruna with the d.c. electric field above Ramfjord, deduced from 'STARE' data, but these data showed no clear correlation. This is not surprising considering that the d.c. electric field has a dynamic

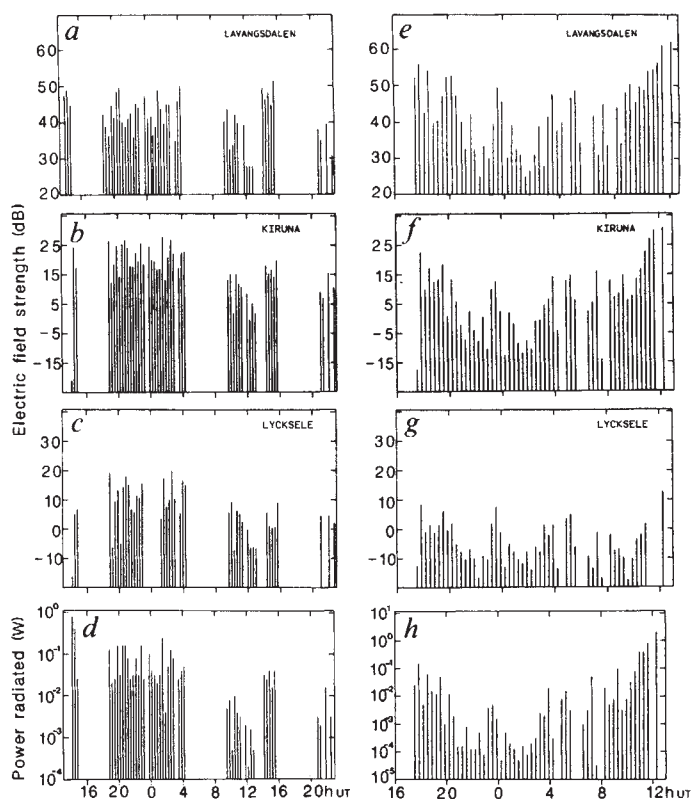


Fig. 2 ELF signals radiated by the polar electrojet antenna. *a-d*, 1.04 kHz; *e-h*, 1.57 kHz. *a, e*, The horizontal r.m.s. magnetic field strength recorded at Lavangsdalen, underneath the source (0 dB = 1 ft). *b, f*, Show the vertical electric field strength recorded at Kiruna (0 dB = $1 \mu\text{V m}^{-1}$ = 3.33 fT). *c, g*, The vertical electric field strength recorded at Lycksele, evaluated from the azimuthal magnetic field components (0 dB = $1 \mu\text{V m}^{-1}$). The equivalent vertical electric field component was presented to assist in the comparison with the Kiruna electric field data. Both electric and magnetic field data were recorded at Lycksele but the converted magnetic field data have been presented here as the magnetic field receivers had the lower noise value and the signals at Lycksele were weak. *d, h*, The effective power radiated by the polar electrojet antenna evaluated from the Kiruna and Lycksele field strength data. Where only the Kiruna data were available at any time, the estimate was made from the Kiruna data alone. These results assume the radiation to be isotropic.

range of $\sim 20\text{--}50 \text{ mV m}^{-1}$ only, whereas the ELF receivers recorded signals varying by more than two orders of magnitude. It seems that a strong electrojet is a necessary but not unique parameter determining the strength of ELF radiation. Interference effects in the ionospheric current source^{3,4} or inhomogeneities in the ionospheric source region may significantly reduce output power even in the presence of a strong electrojet.

Figure 3*a* shows the phase of ELF signals recorded at Kiruna on 1.04 kHz. The large temporal variability of the signal phase is evident and highlights the dangers of using long integration periods when attempting to detect long-path polar electrojet radiation. Also plotted in Fig. 3, where data were available, is the azimuth of the STARE d.c. electric field recorded above Ramfjord. It is evident that there is a significant correlation between the received ELF phase and the azimuth of the d.c. electric field and this is confirmed by the correlation coefficient of 0.88. Rietveld *et al.*¹³ observed clear correlations between d.c. electric fields as measured by STARE and the local ELF signal recorded at Lavangsdalen during a period of enhanced micro-pulsation activity.

Figure 3*b* shows the phase difference, $\Delta\phi$, between signals recorded at Kiruna and Lycksele plotted on the same scale as the Kiruna signal phase. The marked reduction in scatter on this plot clearly shows that the dominant source of signal-phase

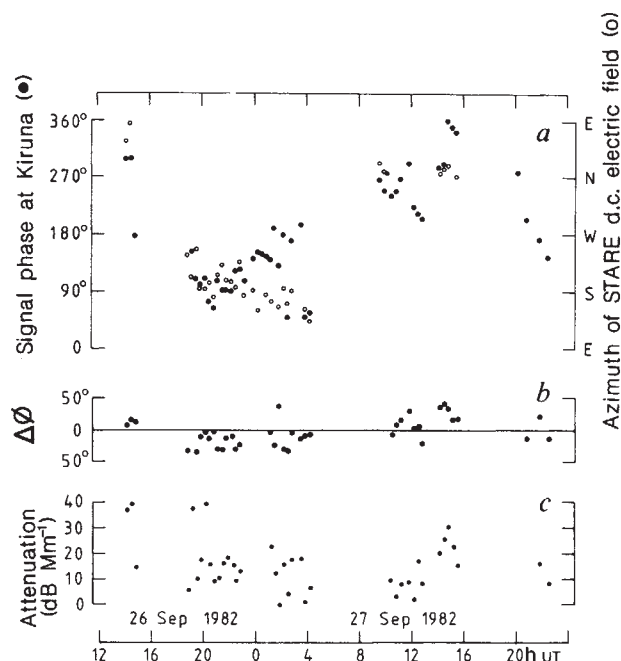


Fig. 3 All data presented in this figure are for a frequency of 1.04 kHz. *a*, The phase of the ELF signal recorded at Kiruna (●), together with the azimuth of the d.c. electric field above Ramfjord (○), as measured by STARE. *b*, The phase difference, $\Delta\phi$, between the ELF signals recorded at Kiruna and Lycksele. *c*, The computed Earth-ionosphere waveguide attenuation constant evaluated from the Kiruna and Lycksele field strength data.

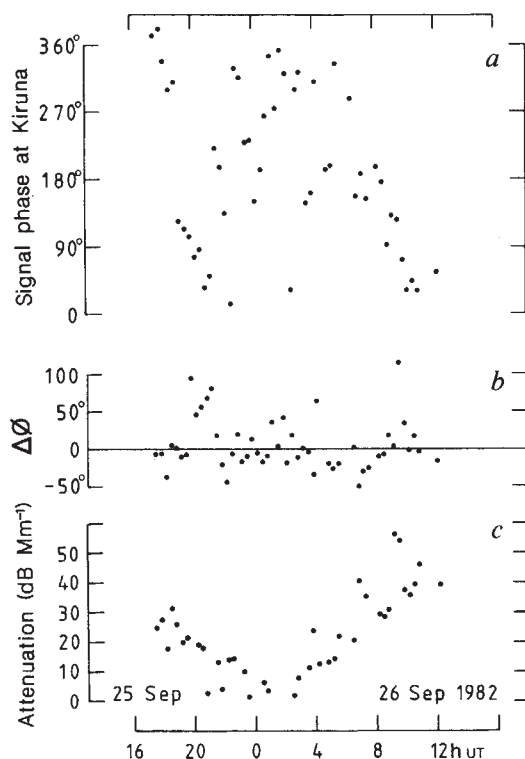


Fig. 4 All data presented in this figure are for a frequency of 1.57 kHz. *a*, The phase of the ELF signal recorded at Kiruna; *b*, the phase difference, $\Delta\phi$, between the ELF signals recorded at Kiruna and Lycksele; *c*, the computed Earth-ionosphere waveguide attenuation constant evaluated from the Kiruna and Lycksele field strength data.

variability is provided by the ELF source mechanism and not by propagation effects. A slight diurnal change in the differenced phase data is just detectable and the 15% change in phase velocity required to produce it is within the range of previous experimental observations.

Figure 3c shows the waveguide attenuation constant computed from the two-station propagation data. The range of values from 5 to 25 dB Mm⁻¹ is not unreasonable for daytime propagation^{10,14-17}, but the larger values of attenuation (>35 dB Mm⁻¹) around 1420 UT and 1930 UT are greater than would be expected. However, inspection of the Kiruna ionosonde data shows that significant auroral precipitation was present during the period of these observations and this may have led to the high attenuation values.

Figure 4 shows a similar data set to Fig. 3 but for 1.57 kHz. (Absence of radar backscatter above Tromsø from 1850 to 0215 UT and technical reasons during the rest of the interval meant that no STARE electric field data were available for comparison with these data.) The signal phase at Kiruna again shows marked variability and that this is predominantly a source effect is evident from the reduced scatter on the Kiruna-Lycksele phase difference data ($\Delta\phi$). Unlike the 1.04-kHz data set, however, the 1.57-kHz attenuation constant exhibits a clear diurnal variation, changing from ~5 dB Mm⁻¹ at night to ~40 dB Mm⁻¹ during the day. The daytime attenuation values from 15 to 40 dB Mm⁻¹ are within the range of previous experimental and theoretical evaluations^{10,14-17}. The low values between 5 and 10 dB Mm⁻¹ recorded just after local midnight are less than might be expected for the zero order mode at 1.57 kHz. In fact, such low attenuation values are not expected until after the 'cut-on' of the first order modes at frequencies above ~1.8 kHz. Note that there was no significant *D*-region absorption on the Kiruna ionosonde records during these nocturnal observations.

The significant correlation between the signals recorded at Kiruna and Lycksele clearly shows that the polar electrojet antenna is an effective source in the Earth-ionosphere waveguide. These data also show the marked temporal variability of the power of the signals emitted, with values peaking at around 1-2 W for frequencies between 1 and 2 kHz. With a power input of 1 MW this represents a power efficiency of ~0.0001%. A long-wire antenna array based on the 'Sanguine' design would radiate 2 kW at a similar frequency with the same input power¹⁸. The parameters of the Earth-ionosphere waveguide derived from these two-station results are in broad agreement with previous data obtained using conventional ELF sources.

We thank the staff of the geophysical observatories at Kiruna and Lycksele for assistance in housing and operating the ELF receivers and the Deutsche Forschungsgemeinschaft (DFG) for its support of the heating project. R.B. thanks the Max-Planck-Institut for financial support and the New Zealand Department of Scientific and Industrial Research for leave to allow him to visit the Max-Planck-Institut für Aeronomie.

Synthesis of silica-sodalite from non-aqueous systems

D. M. Bibby & M. P. Dale*

Chemistry Division, DSIR, Private Bag, Petone, New Zealand

The aluminosilicate sodalite has been known since antiquity¹. It occurs naturally as the polysulphide-containing mineral lapis lazuli¹ and as a salt-bearing feldspathoid² and has also been synthesized as the zeolite basic sodalite with the formula Na₈(Al₆Si₆O₂₄)·xNaOH·(8-2x)H₂O (ref. 2). The structure is an open framework of relatively large cages which are accessible, to water molecules and some ions, through relatively small windows of ~2.1 Å free diameter. Large ions or molecules may be trapped in the cages during synthesis. In aluminosilicate sodalite, the cages contain the number of cations required to compensate for the lattice charge, together with varying amounts of H₂O and salts such as NaOH and NaCl. Aluminosilicate sodalites usually have an Al/Si ratio of about unity although alumina analogues of sodalite have been synthesized³ and a silica-rich sodalite (TMA sodalite) has been prepared from aqueous hydrothermal systems^{4,5} in the presence of tetramethylammonium ions. We report here a new method for the synthesis of sodalite, either in a silica-rich aluminosilicate form or in a novel pure-silica form. Our method of synthesis of sodalite is unusual in that the solvent system is essentially non-aqueous. In the absence of alumina, a new material is produced, namely a pure-silica form of sodalite which, following previously published suggestions⁶, we call silica-sodalite. If the 'solvent' system is ethylene glycol then the unit cell composition approximates to Si₁₂O₂₄·2C₂H₄(OH)₂.

Sodalite is conventionally synthesized either from alkaline aqueous hydrothermal solutions containing sources of active silica and alumina² or by a dry method, involving reaction of mixtures of salts and sources of silica and alumina at high temperature (~800 °C) (ref. 2).

Typical starting materials used here in the non-aqueous preparation of silica-sodalite comprise a solvent system (such as ethylene glycol, propanol), an active form of silica (such as fumed silica) and NaOH, with a typical reactant molar composition of 2SiO₂·3NaOH·40Z, where Z is the solvent used. Silica-sodalite forms from this mixture after 15-25 days at 150 °C. The water generated during the reaction of NaOH with ethylene glycol or silica does not seem to be essential as it can be distilled off at the beginning of the reaction without affecting its progress. However, as no attempt was made to ensure an absolutely water-free system we cannot preclude a (catalytic) role for trace amounts of water.

Incorporation into the reactant mixture of variable amounts of an active form of alumina produces aluminosilicate sodalites with varying compositions, and substantially reduces the nucleation and crystallization times. However, it is possible that the Al is inhomogeneously distributed in this aluminosilicate sodalite, either as Al-rich zones in the sodalite crystals or as crystals with different Al contents.

The morphology of the silica-sodalite is cubic with octahedral and dodecahedral modifications. If the Si concentration of the reactant mixture, or the Al/Si ratio, is increased, the morphology changes, with the octahedral and dodecahedral modifications becoming dominant. The silica-sodalite crystals are ~30 µm across and uniform in size. As the Al content increases, the size becomes less uniform, ranging from ~5 to ~20 µm.

Sodium ions are incorporated into the aluminosilicate sodalite in sufficient quantities to act as counter ions, indicating that Al substitutes for Si in the sodalite lattice. However, analysis of the silica-sodalite showed the sodium content to be essentially zero so that this method of synthesis does not result in occluded NaOH. In silica-sodalite the cages are, nevertheless, occupied. The observed weight loss and composition of the volatiles

* Permanent address: Ceramco Ltd, Private Bag, New Lynn, Auckland, New Zealand.

Received 15 April; accepted 29 May 1985.

- Kapustin, I. N. *et al. J. exp. theor. Phys. Lett.* **25**, 228-231 (1977).
- Stubbe, P., Kopka, H. & Dowden, R. L. *J. geophys. Res.* **86**, 9073-9078 (1981).
- Stubbe, P., Kopka, H., Rietveld, M. T. & Dowden, R. L. *J. atmos. terr. Phys.* **44**, 1123-1135 (1982).
- Ferraro, A. J. *et al. J. atmos. terr. Phys.* **46**, 855-865 (1984).
- Ferraro, A. J. *et al. J. atmos. terr. Phys.* **44**, 1113-1122 (1982).
- Lunnen, R. J., Lee, H. S., Ferraro, A. J., Collins, T. W. & Woodman, R. F. *Nature* **311**, 134-135 (1984).
- Tripathi, V. K., Chang, C. L. & Papadopoulos, K. *Radio Sci.* **17**, 1321-1326 (1982).
- Barr, R. & Stubbe, P. J. *J. atmos. terr. Phys.* **46**, 315-320 (1984).
- Barr, R. & Stubbe, P. *Radio Sci.* **19**, 1111-1122 (1984).
- Chapman, F. W., Jones, D. L., Todd, J. D. W. & Challinor, R. A. *Radio Sci.* **1**, 1273-1282 (1966).
- Wait, J. R. *Electromagnetic Waves in Stratified Media* (Pergamon, New York, 1962).
- Watt, A. D. *Radio Engineering*, 171-394 (Pergamon, New York, 1967).
- Rietveld, M. T., Kopka, H., Nielsen, E., Stubbe, P. & Dowden, R. L. *J. geophys. Res.* **88**, 2140-2146 (1983).
- Challinor, R. A. *J. atmos. terr. Phys.* **29**, 803-810 (1967).
- Galejs, J. *Terrestrial Propagation of Long Electromagnetic Waves* (Pergamon, Oxford, 1972).
- Barr, R. *ELF Radio Wave Propagation* (ed. Holtet, J.) 225-231 (Reidel, Hingham, 1974).
- Challinor, R. A. *J. atmos. terr. Phys.* **29**, 995-1003 (1967).
- Bernstein, S. L. *et al. Proc. IEEE* **62**, 292-312 (1974).

evolved during thermogravimetric analysis/mass spectrometry⁷ of silica-sodalite prepared from ethylene glycol suggest the presence of an ethylene glycol molecule in each sodalite cage. Not all the organic carbon could be removed on calcination as the silica-sodalite became black, and remained so even after extended periods at 1,000 °C in air.

The X-ray powder diffraction patterns of silica-sodalite, of our aluminosilicate sodalite and of conventionally prepared aluminosilicate sodalite show no significant difference. The unit-cell dimension of silica-sodalite, determined from the X-ray powder diffraction pattern, is 8.836 Å, in the typical range for sodalites⁶. Silica-sodalite is, therefore, the zero-aluminium end-member of the aluminosilicate sodalite series.

We have synthesized silica-sodalite from systems containing both ethylene glycol and propanol. Each of these molecules has a molecular volume similar to that of the tetramethylammonium ion trapped in the cages of TMA sodalite^{4,5} and to the K⁺ cluster which has been produced⁸ in the sodalite cage. As we have been unsuccessful in synthesizing silica-sodalite from systems containing a wide range of other organic liquids, we suggest that spatial effects are important.

Silica-sodalite joins the group of low-density polymorphs of silica, such as the naturally occurring melanophlogite⁹, dodecasil-3C (ZSM-39)^{10,11} and other essentially zero-Al-content zeolite endmembers such as silicalite (ZSM-5)^{12,13}, silicalite-2 (ZSM-11)^{14,15} and ZSM-12 (ref. 16). However, unlike these compounds, which are rich in five-membered rings, silica-sodalite contains only four- and six-membered rings. The formation of structures with four- and six-membered rings is generally a characteristic of high-alumina-content framework silicates synthesized from systems containing inorganic bases¹⁷, while five-membered rings are more characteristic of high-silica sili-

cates (SiO₂/Al₂O₃ > 20) synthesized from systems containing organic bases. The formation of silica-sodalite with a structure containing four- and six-membered rings may be a characteristic of reactions in non-aqueous systems and suggests that in the appropriate conditions other framework aluminosilicates could be synthesized in the silica-rich or pure-silica form.

Although water is an exceptional solvent, the reactions of silica in non-aqueous solvents are not necessarily unexpected. Silica will complex with many organic molecules, such as hydroxy- and amine-compounds¹⁸, and so is readily solvated. In principle, therefore, many organic solvents are available for the non-aqueous synthesis of silicates. While we do not necessarily expect to find the range of reactions that occur in water, we suggest that a rich area of silicate and zeolite chemistry awaits further study.

Received 10 June; accepted 10 July 1985.

1. Seel, F. in *Studies in Inorganic Chemistry* Vol. 5 (eds Muller, A. & Krebs, B.) 67–89 (Elsevier, Amsterdam, 1984).
2. Barrer, R. M. *Hydrothermal Chemistry of Zeolites* (Academic, London, 1982).
3. Depmeier, W. *Acta crystallogr.* C40, 226–231 (1984).
4. Baerlocher, Ch. & Meier, W. M. *Helv. chim. Acta* 52, 1853–1860 (1969).
5. Kuhl, G. H. *Inorg. Chem.* 10, 2488–2495 (1971).
6. Henderson, C. M. B. & Taylor, D. S. *Spectrochim. Acta* 33A, 283–290 (1977).
7. Bibby, D. M., Parker, L. M. & Patterson, J. E. *Zeolites* 4, 168–174 (1984).
8. Edwards, P. P. et al. *JCS Chem. Commun.*, 982–984 (1984).
9. Skinner, B. J. & Appleman, D. E. *Am. Miner.* 48, 854–867 (1963).
10. Gies, H., Liebau, F. & Gerke, H. *Angew. Chem. int. Edn* 21, 206–207 (1982).
11. Schlenker, J. L., Dwyer, F. G., Jenkins, E. E., Rohrbach, W. J. & Kokotailo, G. T. *Nature* 294, 340–341 (1981).
12. Flanigen, E. M. et al. *Nature* 271, 512–516 (1978).
13. Dwyer, F. G. & Jenkins, E. E. US Patent 3941871 (1976), Re 29948 (1979).
14. Bibby, D. M., Milestone, N. B. & Aldridge, L. P. *Nature* 280, 664–665 (1979).
15. Kokotailo, G. T., Chu, P., Lawton, S. L. & Meier, W. M. *Nature* 275, 119–120 (1978).
16. Rosinski, E. J. & Rubin, M. K. US Patent 3832449.
17. Flanigen, E. M. in *Proc. 5th Int. Conf. Zeolites* (ed. Rees, L. V.) 760–780 (Heyden, London, 1980).
18. Iler, R. K. *The Chemistry of Silica* (Wiley, New York, 1979).

Quantitative importance of alkalinity flux from the sediments of acid lakes

R. Carignan

INRS-Eau, 2700 Einstein, Ste-Foy, Québec, Canada G1V 4C7

The recognition of lake acidification as an important environmental problem has recently stimulated the development of predictive models^{1–4} linking strong acid loading to lake pH or alkalinity. Among these models, the mass balance approach² has been used to predict lake alkalinity from measured stream and precipitation chemistry and from watershed hydrology. This approach requires that all important positive or negative alkalinity fluxes within the system be considered. The diffusive loss of H⁺ to the sediments and loss of HCO₃[–] from the sediments constitute potentially important sources of alkalinity in acid lakes. However, the relative importance of these fluxes has generally been overlooked and the presently available information appears to be limited to a few studies on the generation of alkalinity in anoxic hypolimnia^{3,4}. The flux of alkalinity across the sediment–water interface is estimated here for one lake subject to relatively high acid deposition. For this lake, the estimated annual diffusive loss of alkalinity from the sediment to the overlying water was found to be as large as the combined measured acidity input from the watershed and from direct precipitation to the lake, thus showing the significant role of sediment–water interactions in the alkalinity balance of some acid lakes.

Clearwater Lake (46° 22' N–81° 03' W; area, 7.65 × 10⁵ m²; mean depth, 8.4 m; maximum depth, 22 m) is located 12 km south-west of Sudbury, Ontario, and 13 km south of the Copper Cliff smelter, a major point source of atmospheric sulphur dioxide emissions⁵. The epilimnion thickness usually increases from 9–12 m in June to 14–16 m in September. Dissolved oxygen is never completely depleted in the hypolimnion but usually decreases to 2–4 parts per million (p.p.m.) in the deepest part

(>20 m) of the basin towards the end of the summer stratification⁶. Because of the regionally high bulk deposition of strong acids, the scarcity of soils in the watershed and the felsic nature of the bedrock, the pH of the surface water is relatively low and has varied between 4.1 and 4.6 since 1973 (ref. 5).

Porewater samples were obtained by *in situ* dialysis^{7,8} using a 1-cm vertical resolution at four stations (designated 10 m North, 10 m South, 15 m South and 20.2 m Central) in late summer 1982 and early summer 1984. The Plexiglass dialysers had two vertical rows of compartments 7.0 cm long × 0.65 cm wide × 0.8 cm deep, spaced 8.3 cm apart (centre to centre) and covered with a Gelman HT-450 or HT-200 membrane. One row of compartments was used for measurements of pH, ΣCO₂ (ref. 8) and ΣH₂S (ref. 9) whereas the other was reserved for the determination of [SO₄^{2–}] by ion chromatography, [Ca], [Mg], [Fe], [Mn] by atomic absorption, and NH₄⁺ by automated colorimetry. The measured porewater metals were assumed to be in the form of the divalent cation. Representative [H⁺] and [HCO₃[–]] profiles are presented in Fig. 1. Most profiles showed steep negative [H⁺] and/or steep positive [HCO₃[–]] gradients within the top 0–2 cm of the sediments. The 10 m South station always showed atypical [H⁺], [Ca²⁺], [Mg²⁺], [ΣCO₂] and [SO₄^{2–}] profiles, apparently resulting from a downward advective transport of porewaters at this particular station¹⁰. Consequently, the results from this latter station were ignored in subsequent calculations. The H⁺ neutralization and HCO₃[–] production observed within the superficial sediments of the three other stations can be attributed to several possible reactions including the microbial reduction of Fe(III), Mn(IV), NO₃[–] and SO₄^{2–} (refs 4, 11), the weathering of aluminosilicates and oxides¹², and ion exchange of H⁺ with other mobile sedimentary cations. In these sediments, however, the reduction of Mn(IV) is probably not important, as porewater [Mn²⁺] never exceeds 6 μmol l^{–1} in the top 10 cm. Moreover, alkalinity generation from NO₃[–] reduction is also probably negligible since it is never found in concentrations exceeding 5 μmol l^{–1} in the superficial porewaters.

Diffusive fluxes are usually calculated according to Fick's first law, as applied to the sedimentary environment¹¹:

$$J = -\phi D_s \frac{dC}{dx} \quad (1)$$

where ϕ is the porosity, D_s is the molecular diffusion coefficient including the effect of sediment tortuosity, and dC/dx is the concentration gradient. In the following flux calculations I have assumed $\phi = 1$ and $D_s = D_o$ (the molecular coefficient of diffusion in water). Both assumptions are justified here by the high porosities (>0.96) measured in the top 3 cm of these sediments. Below pH 6.5, the alkalinity of most natural freshwaters can be expressed as

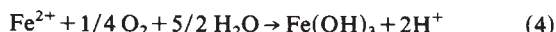
$$\text{ALK} = [\text{HCO}_3^-] + [\text{HS}^-] - [\text{H}^+] \quad (2)$$

In the anoxic porewaters of Clearwater Lake, $[\text{HS}^-]$ rarely exceeds $5 \mu\text{mol l}^{-1}$ and is always negligible compared with $[\text{HCO}_3^-] - [\text{H}^+]$ (ref. 10). Therefore, the flux of alkalinity across the sediment-water interface can be approximated as:

$$J_{\text{ALK}} = J_{\text{HCO}_3^-} - J_{\text{H}^+} = -\phi \left(D_o^{\text{HCO}_3^-} \frac{d[\text{HCO}_3^-]}{dx} - D_o^{\text{H}^+} \frac{d[\text{H}^+]}{dx} \right) \quad (3)$$

Furthermore, the diffusion of ions can be significantly enhanced or reduced by the presence of the electrical potential gradient generated when ions of different chemical mobilities are co-diffusing. This coupling effect was calculated for each species according to Lasaga¹³, assuming no significant ion pairing.

Table 1 shows the observed concentration gradients and the resulting H^+ , Fe^{2+} , Ca^{2+} , Mg^{2+} , NH_4^+ , HCO_3^- and SO_4^{2-} fluxes, calculated on an annual basis. The coupling effect on the fluxes of H^+ , HCO_3^- , NH_4^+ and Fe^{2+} was relatively small (-8 to $+5\%$ of the calculated fluxes without coupling), but important for SO_4^{2-} (10–100%). The average gross alkalinity flux ($J_{\text{HCO}_3^-} - J_{\text{H}^+}$) is thus estimated to be $-0.723 \text{ equiv. m}^{-2} \text{ yr}^{-1}$. However, this flux must be corrected for the H^+ production resulting from the oxidation of the upwardly diffusing $[\text{Fe}^{2+}]$:



Because NH_4^+ and NO_3^- do not accumulate in the overlying water of this lake², the H^+ production resulting from the biological uptake of the diffusing NH_4^+ must also be subtracted from the gross alkalinity flux. The calculated net alkalinity fluxes thus ranged between -0.080 and $-0.440 \text{ equiv. m}^{-2} \text{ yr}^{-1}$, with an average of $-0.207 \text{ equiv. m}^{-2} \text{ yr}^{-1}$. Because of limited replication and large horizontal variability, no obvious depth or seasonal

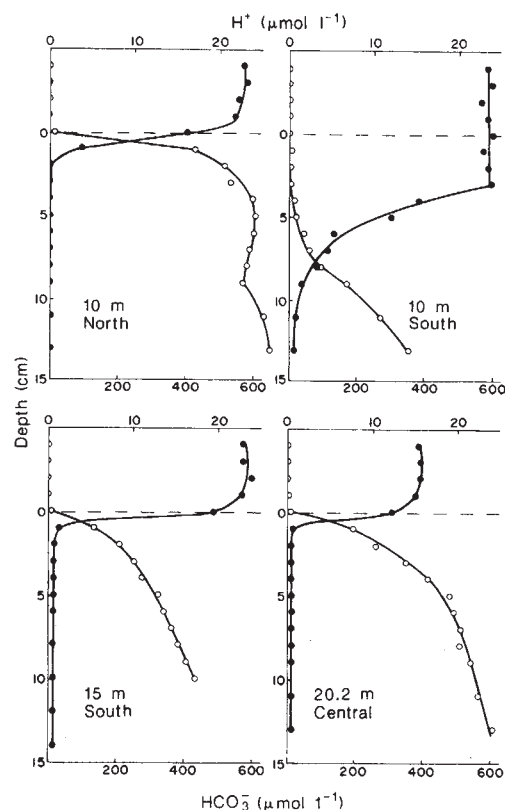


Fig. 1 Porewater profiles of $[\text{H}^+]$ (●) and $[\text{HCO}_3^-]$ (○) obtained on June 21 1984 at four stations in Clearwater Lake. The dashed lines represent the sediment-water interface.

trend could be detected. The net alkalinity flux was similar to the mean H^+ flux ($-0.233 \text{ equiv. m}^{-2} \text{ yr}^{-1}$) alone, indicating that most of the alkalinity flux due to HCO_3^- ($-0.490 \text{ equiv. m}^{-2} \text{ yr}^{-1}$) was offset by an equivalent Fe^{2+} flux ($-0.412 \text{ equiv. m}^{-2} \text{ yr}^{-1}$). Such a result is expected because Fe^{2+} and HCO_3^- are the two major ions in the anoxic porewaters of this lake and must co-diffuse towards the interface. Note that the fluxes of Ca and Mg across the sediment-water are negligible relative to H^+ or HCO_3^- , indicating that ion exchange and weathering reactions are unimportant H^+ sinks in these sediments.

Table 1 Measured concentration gradients of H^+ , Fe^{2+} , Ca^{2+} , Mg^{2+} , NH_4^+ , HCO_3^- and SO_4^{2-} between 0 and 1 cm in the porewaters of three stations in Clearwater Lake

Station	Temperature (°C)	$\frac{d[\text{H}^+]}{dx}$	$\frac{d[\text{Fe}^{2+}]}{dx}$	$\frac{d[\text{Ca}^{2+}]}{dx}$	$\frac{d[\text{Mg}^{2+}]}{dx}$	$\frac{d[\text{NH}_4^+]}{dx}$	$\frac{d[\text{HCO}_3^-]}{dx}$	$\frac{d[\text{SO}_4^{2-}]}{dx}$
				$(\mu\text{mol cm}^{-3} \text{ cm}^{-1})$				
10 m N	13.0	-0.0167	0.197	-0.013	0.008	0.012	0.330	-0.027
21 June '84		(0.0084)	(0.099)	(0.127)	(0.047)	(0.006)	(0.168)	(0.157)
10 m N	13.0	-0.0124	0.339	0.005	-0.002	0.018	0.364	-0.065
21 June '84		(0.0100)	(0.170)	(0.135)	(0.046)	(0.010)	(0.191)	(0.166)
15 m S	11.3	-0.0189	0.031	0.005	0.000	0.017	0.097	-0.014
21 June '84		(0.0104)	(0.016)	(0.130)	(0.049)	(0.019)	(0.054)	(0.174)
15 m S	11.3	-0.0117	0.057	-0.003	0.002	0.020	0.044	-0.018
21 June '84		(0.0080)	(0.041)	(0.130)	(0.047)	(0.020)	(0.030)	(0.176)
15 m S	15.1	-0.0112	0.050	-0.002	0.002	0.015	0.019	-0.023
21 September '82		(0.0060)	(0.028)	(0.143)	(0.056)	(0.027)	(0.010)	(0.194)
20.2 m C	10.9	-0.0085	0.060	-0.020	-0.004	—	0.048	-0.020
21 June '84		(0.0055)	(0.031)	(0.118)	(0.047)	—	(0.031)	(0.163)
20.2 m C	10.9	-0.0116	0.140	-0.018	-0.006	0.023	0.150	-0.036
21 June '84		(0.0065)	(0.070)	(0.123)	(0.048)	(0.013)	(0.081)	(0.159)
20.2 m C	11.9	-0.0002	0.192	0.010	0.00	0.067	0.410	-0.037
21 September '82		(0.0003)	(0.352)	(0.144)	(0.058)	(0.088)	(0.675)	(0.112)
$\bar{J}$ ($\text{mol m}^{-2} \text{ yr}^{-1}$)				0.233	-0.206	-0.003	-0.001	-0.104
±s.e.		0.040	0.057	0.011	0.004	0.028	0.138	0.018

Mean concentrations between 0 and 1 cm, assuming constant gradients within the same depth interval, are shown in parentheses. $[\text{HCO}_3^-]$ was calculated from $[\Sigma\text{CO}_2]$ and the first dissociation constant of H_2CO_3 (ref. 12), at *in situ* temperature. The calculated mean fluxes ($\bar{J}$) and s.e.m. are shown in the last two rows. Self-diffusion coefficients were taken from Li and Gregory¹⁷ and corrected for *in situ* temperature¹⁷.

Equation (3) does not take into consideration possible random transport mechanisms of porewater fluids such as benthic irrigation or gas ebullition. Although gas bubble formation does not occur in the sediments of this lake, the sediment surface was colonized by about 100–400 tube-forming invertebrates per m² at all stations (visual inspection by scuba). Moreover, the concentration gradients may have been underestimated because of the limited vertical resolution (1 cm) of the sampling equipment. The above fluxes should therefore be considered as minimum estimates of the true fluxes.

An estimate of the whole lake alkalinity flux to the sediments can be obtained by multiplying the mean flux for the three stations by the total sediment area of the lake. A visual inspection of the littoral of Clearwater Lake revealed that little sediment accumulation occurs above a depth of about 5 m, giving a total sediment area of 4.98×10^5 m² for this lake. If the mean alkalinity flux found for the 10–20.2-m deep sediments (60% of the total sediment area) is assumed to be also representative of the 5–9-m sediments, a whole lake flux value of -0.99×10^5 equiv. yr⁻¹ ($+0.99 \times 10^5$ equiv. yr⁻¹ when the lake water is chosen as the frame of reference) is obtained. Winter fluxes could be lower because the rates of microbial Fe(III) and SO₄²⁻ reduction are expected to be temperature dependent. Mean sediment temperatures vary from 9 to 16 °C between June and September, and are probably 5–10 °C lower in winter. However, in the absence of empirical data on composite diffusivities (molecular + benthic irrigation) at the interface and on the temperature dependence of alkalinity fluxes for this lake, we may conclude that the minimum estimate of the annual alkalinity flux from the sediment to the lake probably lies between 0.5 and 2×10^5 equiv. yr⁻¹.

The importance of this flux in the annual alkalinity budget of Clearwater Lake can be appreciated when it is compared with its alkalinity inputs from the watershed and direct precipitation. The average measured¹⁴ alkalinity input from the watershed and precipitation to the lake was -0.86×10^5 equiv. yr⁻¹ for the period June 1978 to May 1979. Dillon and Scheider¹⁴ concluded that this value compared favourably with the theoretical input (-0.73×10^5 equiv. yr⁻¹) required to maintain an average observed alkalinity of -50μ equiv. l⁻¹ in this lake. However, this estimate does not take into consideration the possibly important negative alkalinity contribution caused by direct SO₂ deposition on the lake. The direct deposition of SO₄²⁻ precursors to the lake has been estimated¹⁵ to be 0.156 equiv. m⁻² yr⁻¹ (1.19 equiv. yr⁻¹ for the whole lake). As this SO₄²⁻ appears to originate from SO₂ oxidation¹⁵, an additional alkalinity input of -1.19×10^5 equiv. yr⁻¹ must be added to the above 0.86×10^5 equiv. yr⁻¹ value to give a total alkalinity input of -2.05 equiv. yr⁻¹ to the lake. As a result, the estimated total alkalinity input cannot be reconciled with the theoretical input of -0.73×10^5 equiv. yr⁻¹ unless an unaccounted for internal alkalinity source is invoked. From mass balance considerations, the size of this internal input should be of the order of $+1.3 \times 10^5$ equiv. yr⁻¹ and is consistent with the estimated alkalinity flux from the sediment (0.5 – 2×10^5 equiv. yr⁻¹). This clearly shows the necessity for taking sediment–water interactions into consideration when modelling the alkalinity dynamics of some acid lakes. The intensity of the [H⁺] and/or [HCO₃⁻] gradients across the sediment–water interface, and the importance of the resulting fluxes will depend both on the [H⁺] in the overlying waters and on the presence of an efficient H⁺ sink and/or HCO₃⁻ source located close to the interface. Fe(III) reduction, combined with SO₄²⁻ reduction and the permanent immobilization of the resulting Fe²⁺ as FeS according to the reaction:



probably constitutes an important H⁺ sink/HCO₃⁻ source in the sediments of Clearwater Lake. The importance of such a reaction in these sediments is supported by the abundance (2–5%) of Fe(III) at the sediment surface¹⁶, by the presence of a pronounced SO₄²⁻ flux at the interface (Table 1), by the formation of FeS 1–4 cm below the interface¹⁰ and by the observed agreement

between the calculated net alkalinity and SO₄²⁻ fluxes (Table 1) and the stoichiometry of equation (5). This shows the necessity for considering both SO₄²⁻ and Fe(III) reduction simultaneously when studying the transfer of alkalinity from anoxic sediments to oxic waters. A recent estimate (P. J. Dillon, personal communication) of Fe retention by the lake suggests that only 25% of the Fe required from equation (5) is deposited annually in its sediments. This may indicate that Fe(III) is presently being reduced and buried faster than it is supplied to the sediments, and/or that other sulphide immobilization mechanisms are operative in the surficial sediments. Note that the estimated SO₄²⁻ flux to the sediments of this lake represents only 5–15% of the total load to the lake. Although important in terms of alkalinity generation in acid lakes, this relatively small SO₄²⁻ sink may not be easily detectable from whole lake SO₄²⁻ budgets.

According to equation (5), the alkalinity flux from the sediments to the oxic waters of acid lakes should be expected to increase with increasing supply of any of the components on the left side of the equation, provided that the three others are not limiting. In Clearwater Lake, the complete depletion of SO₄²⁻ observed within the first 4–6 cm of the sediments¹⁰ suggests that equation (5) is presently SO₄²⁻ limited. These results indicate that acid lakes may recover faster than previously thought from recently proposed reductions in SO₂ emissions and emphasize the need for further studies on sulphur retention and diagenesis in these lakes.

This study was partially supported by a NSERC University Research Fellowship. I thank P. J. Dillon, P. G. Campbell and A. Tessier for critical reading of the manuscript.

Received 9 October 1984; accepted 22 July 1985.

1. Wright, R. F. *Acid Rain Research: Predicting Acidification of North American Lakes* (Rep. 4/1983, Norwegian Institute for Water Research, Oslo, 1983).
2. Dillon, P. J. in *Environmental Impacts of Smelters* (ed. Nriagu, J. O.) 283–347 (Wiley, New York, 1984).
3. Schindler, D. W., Wagemann, R., Cook, R. B., Rusczyński, T. & Prokopowich, J. *Can. J. Fish. Aquat. Sci.* 37, 342–354 (1980).
4. Kelly, C. A., Rudd, J. W. M., Cook, R. B. & Schindler, D. W. *Limnol. Oceanogr.* 27, 868–883 (1982).
5. Lazerte, B. D. & Dillon, P. J. *Can. J. Fish. Aquat. Sci.* 41, 1664–1677 (1984).
6. Yan, N. D. & Miller, G. E. in *Environmental Impacts of Smelters* (ed. Nriagu, J. O.) 243–282 (Wiley, New York, 1984).
7. Hesslein, R. H. *Limnol. Oceanogr.* 21, 912–914 (1976).
8. Carignan, R. *Limnol. Oceanogr.* 29, 667–670 (1984).
9. Cline, J. D. *Limnol. Oceanogr.* 14, 454–458 (1969).
10. Carignan, R. & Nriagu, J. O. *Geochim. cosmochim. Acta* (in the press).
11. Berner, R. A. *Early Diagenesis* (Princeton University Press, 1980).
12. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* 2nd edn (Wiley, New York, 1981).
13. Lasaga, A. C. *Am. J. Sci.* 279, 324–346 (1979).
14. Dillon, P. J. & Scheider, W. A. in *Modeling of Total Acid Precipitation Impacts* (ed. Schoor, L.) 121–154 (Butterworth, London, 1984).
15. Dillon, P. J., Jeffries, D. S. & Scheider, W. A. *Wat. Air Soil Pollut.* 18, 241–258 (1982).
16. Tessier, A., Rapin, F. & Carignan, R. *Geochim. cosmochim. Acta* 49, 183–194 (1985).
17. Li, Y.-H. & Gregory, S. *Geochim. cosmochim. Acta* 38, 703–714 (1974).

Tree ring D/H ratio from Kenya, East Africa and its palaeoclimatic significance

R. V. Krishnamurthy & S. Epstein

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

The D/H ratio of non-exchangeable hydrogen atoms in tree rings is a reliable record of the stable isotope ratio of source water used by a given tree during its growth^{1–6}. Depending on the site of growth, this source water has been shown to be the summer rains or groundwater or both^{7–9}. Because the isotope ratio of any of these source waters is influenced by climate, tree ring isotope data provide a fine climatic signature in most cases. However, isotope dendroclimate studies so far have been focused mainly in North America^{2,4,7,9,10} and to a lesser degree in Europe. We report here analysis of a record spanning 145 yr from a tree, *Juniperus procera*, from Kenya in East Africa. The D/H ratio correlates reasonably well with the recorded levels of the adjacent Lake Victoria and

brings out interesting palaeoclimatic inferences from a region for which very little such data are available. Significantly, the relatively severe arid conditions around 1920s and the record rainfall around 1961 are reflected in the tree ring isotope ratio. This suggests that such tree ring isotope records could be used in the palaeoclimatic reconstruction of this area and could be applied to tropical environments.

The tree was sampled in 1979 and comes from the Mau Narok forest (0°40' S, 36° E) at an altitude of 2,700 m. The D/H ratio of cellulose nitrate, which represents the non-exchangeable hydrogen atoms, prepared from 5-yr increments of the raw wood was measured according to standard procedures^{3,11}. The δD values of the rings are plotted as a function of time in Fig. 1a. $\delta D = [(D/H)_{\text{sample}} / (D/H)_{\text{SMOW}} - 1] \times 10^3$, where SMOW is the standard mean ocean water. The overall analytical precision for D/H measurements is $\pm 2\%$.

The δD values are generally higher than that reported for North American and other mid- and high-latitude tree ring samples. This arises from the fact that the δD in precipitation itself is higher in equatorial regions compared with that in mid and high latitudes, the so-called "latitude effect"¹². Of particular interest are the δD values of our samples covering the period 1918–33, which are among the highest found for higher plants.

Previous studies¹, based on the analysis of much plant material belonging to more than 20 different species of terrestrial and aquatic types, have established the relationship to be approximately $\delta D_{\text{water}} = \delta D_{\text{cell.nit}} + 22$, for the source water used by the plant. Using this relationship for the tree ring δD of the above period, the inferred δD values for the soil water of corresponding years turn out to be near +30‰. Such high D/H ratios are reasonable for the region judging by previous work on modern East African meteoric and lake waters. For example, Cerling *et al.*¹³ found the δD of water from Olduvai Gorge and Lake Turkana to range from +4.9 to +26‰ with respect to SMOW and that of Lakes Turkana and Ndutu to be +37.1 and +41.7‰ with respect to SMOW, respectively. Craig¹⁴ also observed unusually heavy values for East African lake waters.

Also shown in Fig. 1b are the recorded annual variations in the level of Lake Victoria for the period 1899–1974 (ref. 15). Lake Victoria, extending to 70,000 km² is the largest water body, ~150 km west of the sample location and also the most commonly used climatic indicator of all East African lakes¹⁶. Because a high proportion of water feeding this lake is derived from direct precipitation over its catchment area, the fluctuations in the levels of the lake have been found to respond sensitively to changes in monthly rainfall. Earlier studies have even attempted to find a correlation between sunspot numbers and the levels of Lake Victoria. According to Butzer¹⁷, the fluctuations in the levels of Lake Victoria over the last century have been a direct response to secular trends of equatorial rains over the Victoria catchment, between latitudes 2° N and 3° S.

In correlating the δD values of the tree rings and the Victoria lake levels, it is important to consider that the tree δD values represent 5-yr averages, whereas the Lake Victoria data are based on monthly readings. Consequently, there might be variations in the lake levels not apparent in the isotope data. Also, the dating of our tree samples depends on the recognition of double rings or missing rings. This is difficult to do in view of the narrow widths of some of the rings, which were as little as 0.5 mm. Taking into account such possible errors in dating the tree rings, there is a generally good correlation between the two records. This inverse correlation, higher strands being marked by lower δD for the tree rings of corresponding period and vice versa, is most noticeable around 1959–63 and 1919–23. The period around 1959–63, as shown by our relatively lower δD values and the elevated Lake Victoria levels, represent one of heavy rainfall in the region. Independent meteorological data do indicate unusually heavy rains in Kenya in 1961 almost three times the average received during earlier corresponding periods¹⁸. The period from 1919 to 1923 coincides with the lowest level registered by Lake Victoria in its recorded history, and seems to be one of very sparse rainfall. In fact, as indicated by

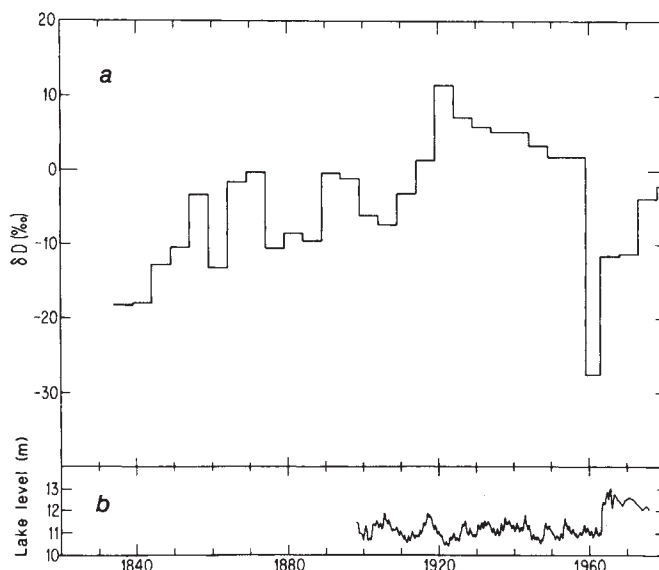


Fig. 1 a, The δD record of the isotopically non-exchangeable hydrogen of cellulose in 5-yr segments covering the time span AD 1834–1979 of a *Juniperus procera* from Kenya, East Africa. b, Variations of the level of Lake Victoria for the period AD 1899–1974 based on gauge readings at Jinja (see ref. 15).

the relatively higher δD values, arid conditions have persisted in varying degrees since the 1920s. Although recorded climatic information of any other type for this time period is absent, it is interesting that visitors to Kenya in the early 1930s mentioned severe droughts¹⁹. On a broader scale, using spatial differences of average pressure levels, Lamb¹⁵ has suggested that the decades between about 1900 and 1953 were the time of strongest general circulation leading to an increased size of the tropical anticyclones and resulting in aridity in the tropical and equatorial zones. That the period since 1968 has been more drought-like in many parts of upper Africa, especially the Sahel zone, is corroborated by recent studies²⁰.

Our tree ring isotope record permits us to draw inferences on the Lake Victoria's levels before 1899, when gauge readings began to be made. Obviously, there have been oscillations more pronounced than those during its recorded period, except again for the steep rise in the 1960s. Significantly, in the 1830s the lake level might have been substantially higher. It would be interesting to see whether this improved condition corresponded to the world-wide culmination of the Little Ice Age¹⁵.

Finally, a striking pattern that is evident while looking at the 145-year record of δD values is the definite trend towards increasing aridity in the region, punctuated only by an anomalously wetter regime in early 1960. Whether the beginning of this century marked the onset of severe desertification in the region is speculative, given our preliminary studies, but more work is clearly warranted. The present study has illustrated the potential of tree ring δD as a sensitive palaeoclimatic indicator in an equatorial situation and future work can reveal not only long and short periodicities in regional climatic regimes, but allow important phenomenon such as monsoon systems, associated with equatorial belts, to be modelled more rigorously.

This research was supported by grant NSF ATM 80/8830 which is gratefully acknowledged. This work was made possible by T. E. Cerling who provided the samples. We also thank Joseph Ruth and Eleanor Dent for technical assistance. Contribution no. 4024 publications of the Division of Geological and Planetary Sciences, California Institute of Technology.

Received 10 April; accepted 4 July 1985.

1. Epstein, S., Thompson, P. & Yapp, C. J. *Science* **198**, 1209–1215 (1977).
2. Epstein, S. & Yapp, C. J. *Earth planet. Sci. Lett.* **30**, 252–261 (1977).
3. Epstein, S., Yapp, C. J. & Hall, J. H. *Earth planet. Sci. Lett.* **30**, 241–251 (1976).
4. Yapp, C. J. & Epstein, S. *Earth planet. Sci. Lett.* **34**, 333–350 (1977).
5. Epstein, S. & Yapp, C. J. *Nature* **266**, 477–478 (1977).
6. Yapp, C. J. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 955–965 (1982).
7. Yapp, C. J. & Epstein, S. *Nature* **297**, 636–639 (1982).

8. White, J. C., Cook, E. R., Lawrence, R. J. & Broecker, W. S. *Geochim. cosmochim. Acta* **49**, 237-246 (1985).
9. Yapp, C. J. & Epstein, S. J. *Geophys. Res.* **90**, No. D2 3747-3752 (1985).
10. Gray, J. & Song, J. S. *Earth planet. Sci. Lett.* **70**, 129-138 (1984).
11. Deniro, M. J. *Earth planet. Sci. Lett.* **54**, 177-185 (1981).
12. Dansgaard, W. *Tellus* **16**, 436-468 (1964).
13. Cerling, T. E., Hay, R. L. & O'Neill, J. R. *Nature* **267**, 137-138 (1977).
14. Craig, H. *Science* **133**, 1702-1703 (1961).
15. Lamb, H. H. *Climate, History and the Modern World* (Methuen, London, 1982).
16. Lamb, H. H. *Geogr. J.* **132**, 183-212 (1966).
17. Butzer, K. W. *Recent history of an Ethiopian Delta* (University of Chicago Press, 1971).
18. Thompson, B. W. & Morth, H. T. *Weather* **20**, 226-227 (1965).
19. Champion, A. J. *Geogr. J.* **89**, 97-118 (1937).
20. *Environmental change in the West African Sahel* (National Academy Press, Washington, 1984).

Induction of functional retinal projections to the somatosensory system

Douglas O. Frost & Christine Metin*

Section of Neuronanatomy, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Optic axons can be induced to form permanent, retinotopic connections in the auditory (medial geniculate, MG) and somatosensory (ventrobasal, VB) nuclei of the Syrian hamster thalamus; this occurs when the principal targets of retinofugal axons are ablated in newborn hamsters and alternative terminal space is created by partial deafferentation of MG or VB¹⁻³. The experimentally induced retinal projection to the somatosensory nucleus occurs by the stabilization of an early, normally transient projection⁴⁻⁵. The present study was undertaken to determine whether the anomalous, stabilized retino-VB projection is functional. Newborn hamsters were operated on to produce permanent retino-VB projections and when the animals were mature, neurophysiological recordings were made in the cortical targets of VB, the first and second somatosensory cortices (SI and SII, respectively). Visual stimulation within well-defined receptive fields reliably evoked multi-unit responses in SI and SII of operated, but not normal hamsters. The representations of the visual field in SI and SII showed a partially retinotopic organization. These results demonstrate that optic tract axons can form functional synapses in the thalamic somatosensory nucleus, and suggest that neural structures which normally process information specific to one sensory modality have the potential to mediate function for other modalities.

Ablations of two of the principal targets of retinofugal axons were made in newborn hamsters anaesthetized by hypothermia:

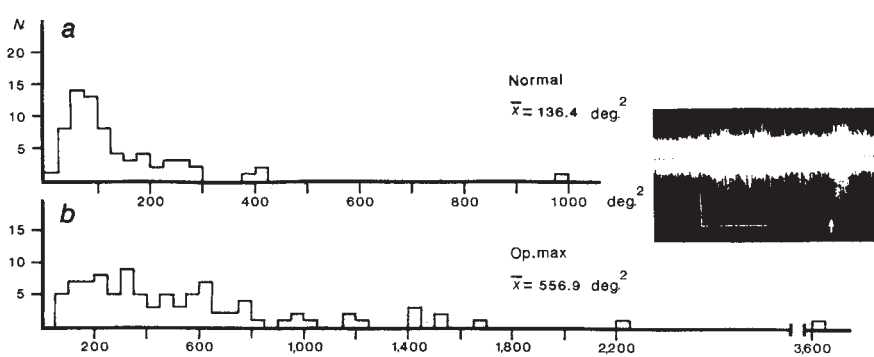
bilateral heat lesions were made in the superior colliculus (SC). The dorsal lateral geniculate nucleus (LGd) was ablated unilaterally by making a heat lesion of the ipsilateral occipital cortex (because of the cortical lesion, the LGd undergoes rapid, retrograde degeneration⁶). The ascending somatosensory afferents to the ipsilateral VB were cut by midbrain hemisection at a mid-SC level; this surgery produces permanent retinal projections to the deafferented VB^{1,2}. Multi-unit recordings were made from the cortical targets of VB, namely SI and SII, in 2 normal control hamsters and 15 operated hamsters, all at least 2.5 months old. As in both normal and operated hamsters retinofugal axons terminate only in the diencephalon or mesencephalon^{1,2,5,7}, the strategy of recording in the cortex ensures that the recordings are from cortical axons of VB neurones or from cortical neurones, that is, postsynaptic to retinofugal axons.

We recorded clear, visually evoked, multi-unit responses (inset, Fig. 1) in SI or SII of all 15 operated hamsters (66 of 82 radially oriented penetrations). Visual responses were strongest and most frequent at 400-600 μm and at 800-1,000 μm below the cortical surface. These depths correspond to the bimodal distribution of thalamocortical afferents originating in VB in normal adult mice⁸ and in adult hamsters with abnormal retino-VB projections². Visual responses in SI and SII varied in strength, and were occasionally comparable to those recorded in primary visual cortex (area 17) in normal hamsters. By contrast, we were unable to record any visually evoked responses in 27 radial penetrations in the somatosensory cortices of normal hamsters, although gentle mechanical stimulation of the vibrissae or body surface evoked brisk responses.

Most of the multi-unit visual receptive fields (RFs) in SI and SII had a single, clearly defined zone of maximal response, although some had two such zones (Figs 2a,b). The mean area of the zones of maximal response in SI and SII (556.9 deg², Fig. 1b) is about 4.1 times the mean size of visual RFs recorded by others⁹ and by us in area 17 of four normal hamsters (Fig. 1a); this difference in RF size is highly significant (Student's *t*-test, $P < 0.005$). There was no systematic variation in the area of the maximal response zones with RF eccentricity in the visual field. The zones of maximal response were sometimes adjoined by a zone of weaker response and in the case of RFs with two zones of maximal response, a zone of weaker response usually linked them to form an elongated RF (Fig. 2b).

* Present address: Laboratoire des Neurosciences de la Vision, Université Pierre et Marie Curie—Paris VI, 4 place Jussieu, Batiment C, 75230 Paris Cedex, France.

Fig. 1 Distributions of RF size. *a*, Areas of 69 RFs recorded from 26 penetrations in area 17 of four normal hamsters. Recording techniques were identical to those used in SI and SII of operated hamsters. *b*, Areas of 93 zones of maximum response from RFs recorded in SI or SII in 15 operated hamsters (Op. max); included are 55 RFs with a single zone of maximum response and 19 RFs with two zones of maximum response. The data from the two types of RF were pooled, since a Student's *t*-test showed no significant difference ($P > 0.6$) between the areas of their maximum response zones. Inset shows a typical, multi-unit, visually evoked response in SII of an adult hamster operated on at birth to produce permanent retino-VB projections. The summed response to 10 diffuse flashes presented at a rate of 2/3 s⁻¹ is shown. The oscilloscope was triggered by stimulus onset; latency of response (arrow) was 160 ms. Horizontal calibration bar, 60 ms; vertical calibration bar, 30 μV .



The summed response to 10 diffuse flashes presented at a rate of 2/3 s⁻¹ is shown. The oscilloscope was triggered by stimulus onset; latency of response (arrow) was 160 ms. Horizontal calibration bar, 60 ms; vertical calibration bar, 30 μV .

Methods. 12 h before recording hamsters were given dexamethasone (0.8 mg intraperitoneal, i.p.). For the recording they were anaesthetized with urethane (210 mg per 100 g i.p.) administered in 3-4 doses over 1.5-2 h. Additional drugs administered included atropine (0.1 mg subcutaneous, s.c.) and dexamethasone (0.8 mg s.c.). Body temperature was maintained at 37 °C using a thermostatically controlled heating pad. Dextrose (5%) in Ringer's (1.0 ml h⁻¹) was administered via a cannula in the femoral vein. In some experiments, the hamsters were paralysed with Flaxedil (20 mg per kg intravenous (i.v.) loading dose; 35 mg per kg h⁻¹, by i.v. infusion) and artificially ventilated. The pupils were dilated with atropine and the cornea protected from drying by non-refractive contact lenses. Visual stimuli were diffuse, stroboscopic flashes, or moving or stationary bars and spots of light back-projected onto a translucent, hemispheric dome centred on the optic disk of the eye contralateral to the recording site. Tungsten microelectrodes with large (30-40 μm long, 2-3 μm diameter, 0.9 M Ω impedance at 1 kHz), uninsulated tips were lowered through the intact dura and into the parietal cortex in a plane perpendicular to the cortical surface. The parietal cortex was then systematically explored for visually responsive regions.

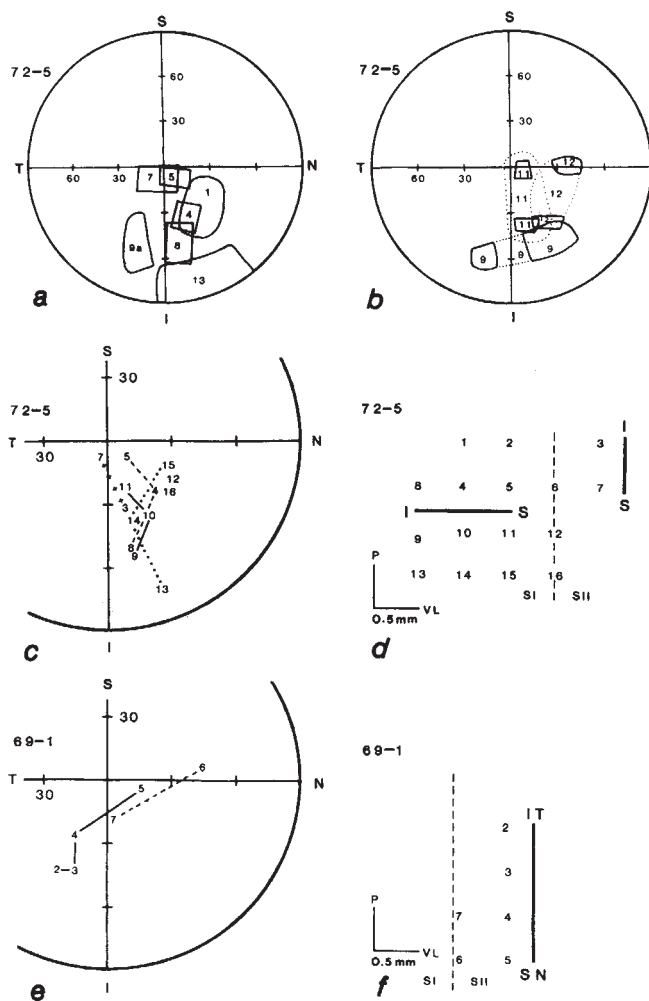


Fig. 2 Multi-unit visual RFs recorded in SI and SII from two adult hamsters operated on day 0 to produce retino-VB projections. RFs were mapped using bars and spots of light projected onto the dome (see Fig. 1 legend). Intersection of the horizontal and vertical axes indicates the projection of the optic disk onto the dome; S, N, I and T represent, respectively, the superior, nasal, inferior and temporal poles of the visual field. The numbers 30 and 60 indicate degrees of eccentricity in the polar coordinate system. *a-d*, Hamster 72-5; *e-f*, hamster 69-1. *a*, RFs with a single zone of maximum responsiveness; only the first RF encountered in each penetration is shown, since as in normal hamster visual cortex, RF position varies somewhat along a vertical electrode penetration. *b*, RFs with two zones of maximum responsiveness (solid outlines) connected by a zone of weaker responsiveness (dotted outlines); corresponding numbers indicate parts of the same RF. In *a* and *b*, for simplicity, RFs are not shown for every penetration. RF 9a in *a* was recorded in the same penetration as RF 9 in *b*, but more superficially. *c*, The positions of RF centres for three medial-to-lateral rows of electrode penetrations in SI (8-4-5, 9-10-11 and 13-14-15) and one posterior-to-anterior row in SII (3-7). RFs of penetrations in the same row are connected. The RFs recorded in penetrations 12 and 16 are not connected to the others as those penetrations were right on the SI/SII border, and are not certain to be part of the SI map. For clarity, the part of the visual field containing RFs is magnified, while the rest is not shown; other conventions as in *a*, *b*. *d*, Positions of all electrode penetrations in the parietal cortex of case 72-5; all but 2 and 6 yielded clear visual RFs. Heavy lines indicate the polarities of the inferior (I)-superior (S) visual axis representations in SI and SII. Broken line indicates the SI/SII border. Scale bars, 0.5 mm. P and VL indicate the posterior and ventrolateral directions respectively. *e*, Positions of RF centres for two posterior-to-anterior rows of electrode penetrations in SII: 2-3-4-5 and 7-6. Conventions as in *c*, *f*, Positions of all electrode penetrations yielding visual RFs in the parietal cortex of case 69-1. Heavy line indicates the polarity of the inferotemporal (IT)-superonasal (SN) visual axis representation in SII. Scale bars and axes as for *d*.

The visual RFs in the somatosensory cortices were preferentially distributed in the lower visual field (Fig. 2); this is consistent with anatomical data showing that the representation of the retina within VB is often incomplete, with the upper retina being preferentially represented¹. In seven cases, we mapped the visual field representations in SI or SII to assess the extent of retinotopic order. In SI and SII the lower-to-upper (or lower temporal-to-upper nasal) visual axis was represented along the medial-to-lateral and posterior-to-anterior directions, respectively (Fig. 2c-f). The visual field representations in SI and SII are oriented with respect to each other in the same way as are the normal somatic representations in SI and SII^{10,11}. There was no obvious representation of the temporal-to-nasal (or upper temporal-to-lower nasal) axis in SI or SII. (The orderly cortical representation of only one visual field axis and the existence of elongated cortical RFs are both probably the result of the way in which the visual field¹ and body¹² maps in VB intersect: retinal afferents representing single points in the visual field are distributed along narrow cylinders of tissue in VB, with adjacent points represented along adjacent cylinders¹. The somatic map in VB consists of a complex, curvilinear array of 'isorepresentation lines', each representing a single body part and projecting beneath a restricted locus of the somatosensory cortical surface¹³. The isorepresentation lines are oriented in such a way that many are cut by multiple, adjoining cylinders representing a sequence of points in the visual field. When afferents driven from a sequence of visual points converge on a single isorepresentation line, the RF recorded at the corresponding cortical locus is elongated along that visual axis whose VB representation is parallel to the isorepresentation line. The preferential orientation of the isorepresentation lines parallel to one axis of the visual field map in VB could also prevent the orderly representation of that axis in the cortex.)

Subsequent to neonatal midbrain hemisections, ascending somatosensory afferents grow rostral to the level of the cut and invade VB despite the presence of the abnormal retino-VB projection (D.O.F., unpublished results). Consistent with this observation, in 60 of 82 penetrations in SI and SII of operated hamsters, we recorded multi-unit responses evoked by somatosensory stimuli. In 48 of the 60 penetrations, both somatosensory and visual responses were recorded. The somatosensory responses were often recorded at the same sites as visual ones, or within 60–100 μ m of sites where purely visual responses were recorded. In some operated hamsters we determined the polarity of the somatic representations; these corresponded to the polarities of the normal somatic maps in SI or SII^{10,11}. As the recording sites were localized in SI or SII, respectively, by histology (see below), the somatotopic organization of the regrown somatosensory projection to VB must be approximately normal. Where both visual and somatosensory responses were recorded, the somatosensory RF almost always constituted one or more mystacial vibrissae or was on the head, neck, shoulders of upper body. This is consistent with the observation that the retino-VB projection is located principally on the lateral aspect of VB¹, which receives input from these regions¹².

The sites from which we recorded visually evoked responses are attributed to SI and SII on the basis of two independent criteria. First, 28 sites were marked with electrolytic lesions made through the recording electrode (Fig. 3). Subsequent histological examination revealed that by cytoarchitectonic criteria¹⁴, these recording sites, as well as others, were in SI or SII. Second, the visual responses were recorded within regions whose somatic representation had the polarity characteristic of SI or SII. Although there are cortical regions from which both somatosensory and visual responses can be recorded in normal rodents, these regions are caudal to the sites from which we recorded in the hamster¹⁵, have a different cytoarchitecture¹⁴, receive their thalamic input from a different, normally multimodal region of the thalamus¹⁶, and unlike SI and SII, do not contain an orderly representation of the body¹⁵.

A long-held tenet of sensory system organization is that in



Fig. 3 Lesion at recording site of the RF from penetration 12 that is illustrated in Fig. 2b. Roman numerals on the left denote cortical layers. The asterisk at the cortical surface indicates the SI/SII border, as determined by the criteria of Caviness¹⁴; in SII there is an overall reduction in cell packing density compared with SI, and the granule cells of layer IV in SII are more widely spaced and globular than in SI. The lesion is located in layer Vb at the SI/SII border. The arrow indicates the lesion; the arrowhead indicates the electrode track; V, blood vessel. Calibration bar, 500 μ m.

adult animals, the primary ascending pathways for the various sensory modalities are segregated in modality-specific, multisynaptic 'labelled lines'¹⁷. However, several instances are now known in which different functional systems are transiently connected early in development^{5,18-23}. When normally transient optic axons in the main thalamic somatosensory nucleus are experimentally stabilized, they form electron microscopically identifiable synaptic connections there²⁴; the stabilized projection also transports intraocularly injected proline via VB to the somatosensory cortex². Here we have shown these connections to be functional in that neurones of the somatosensory nucleus can be activated by visual stimuli within well-defined receptive fields. Thus, neurones from different sensory systems that are normally not connected in adult animals, are capable of forming functional connections in appropriate conditions. The normal elimination of transient inter-system connections must be explained on some basis other than a fundamental incompatibility of the sets of neurones concerned.

The neuronal circuits of the various sensory systems share many morphological and functional features²⁵. This fact and the existence of clear multi-unit visual RFs in the somatosensory cortices of operated hamsters suggest that some of the sensory systems may have enough in common to allow them to process inputs from a modality other than that which they normally mediate. In particular, when supplied with visual input, the circuitry of VB and the somatosensory cortex may permit neurones in these regions to elaborate visual RFs similar to those found in the normal hamster LGd and visual cortex, respectively. We are presently exploring this possibility by means of neurophysiological studies of single visual neurones in VB, SI and SII. Behavioural data confirm the functional competence of other anomalous olfactory²⁶ and visual²⁷ projections in hamsters. Thus, the abnormal visual projections to the somatosensory system may be able to mediate some aspects of visual perception.

This research was supported by NIH grant R01-EY03465 and Basil O'Connor Starter grant 5-417 from the March of Dimes Birth Defects Foundation. C.M. received travel funds from the Fondation pour la Recherche Medicale.

Received 17 April; accepted 4 July 1985.

1. Frost, D. O. *J. comp. Neurol.* **203**, 227-256 (1981).
2. Frost, D. O. *Dev. Brain Res.* **3**, 627-635 (1982).
3. Schneider, G. E. *Brain behav. Evol.* **8**, 73-109 (1973).
4. Frost, D. O. *Neurosci. Abstr.* **9**, 26 (1983).
5. Frost, D. O. *J. comp. Neurol.* **230**, 576-592 (1984).
6. Gudden, B. von Albrecht v. Graefes *Arch. Ophthalm.* **21**, 199-204 (1875).
7. Frost, D. O., So, K.-F. & Schneider, G. E. *Neuroscience* **4**, 1649-1677 (1979).
8. Frost, D. O. & Caviness, V. S. *J. comp. Neurol.* **194**, 369-394 (1980).
9. Tiao, Y.-C. & Blakemore, C. *J. comp. Neurol.* **168**, 459-482 (1976).
10. Sur, M., Nelson, R. J. & Kaas, J. H. *J. comp. Neurol.* **179**, 425-450 (1978).
11. Nelson, R. J., Sur, M. & Kaas, J. H. *J. comp. Neurol.* **184**, 473-490 (1979).
12. Waite, P. M. E. *J. Physiol., Lond.* **228**, 527-540 (1973).
13. Saporta, S. & Kruger, L. *J. comp. Neurol.* **174**, 187-208 (1977).
14. Caviness, V. S. *J. comp. Neurol.* **164**, 247-264 (1975).
15. Wager, E., Mangini, N. J. & Pearlman, A. L. *J. comp. Neurol.* **193**, 187-202 (1980).
16. Caviness, V. S. Jr & Frost, D. O. *J. comp. Neurol.* **194**, 335-368 (1980).
17. Mountcastle, V. B. in *Medical Physiology* 14th edn, Vol. 1 (ed. Mountcastle, V. B.) 327-347 (C. V. Mosby Co., St Louis, 1980).
18. Cucchiari, J. B. & Guillery, R. W. *Neurosci. Abstr.* **6**, 662 (1980).
19. Distel, H. & Hollander, H. *J. comp. Neurol.* **192**, 505-518 (1980).
20. Stanfield, B. B., O'Leary, D. D. M. & Fricks, C. *Nature* **298**, 371-373 (1982).
21. Adams, C. E., Mihailoff, G. A. & Woodward, D. J. *Neurosci. Lett.* **36**, 243-248 (1983).
22. Tolbert, D. L. & Panneton, W. M. *J. comp. Neurol.* **221**, 216-228 (1983).
23. Innocenti, G. M. & Clarke, S. *Dev. Brain Res.* **14**, 143-148 (1984).
24. Campbell, G. & Frost, D. O. *Neurosci. Abstr.* **11** (in the press).
25. Shepherd, G. M. *The Synaptic Organization of the Brain—An Introduction* (Oxford University Press, New York, 1974).
26. Devor, M. *Science* **190**, 998-1000 (1975).
27. Schneider, G. E. *Neuropsychology* **17**, 557-583 (1979).

Monoclonal antibodies demonstrate protection of polymorphonuclear leukocytes against complement attack

A. K. Campbell & B. P. Morgan

Department of Medical Biochemistry,
University of Wales College of Medicine, Heath Park,
Cardiff CF4 4XN, UK

Studies on erythrocytes have shown that the formation of the membrane attack complex on a cell surface inevitably results in lysis. However, it is known that nucleated cells are much more difficult to kill with complement¹⁻³, although the molecular basis of this resistance has never been established. We have shown that a very early intracellular event, occurring within seconds of formation of the attack complex in the membrane, is a rise in cytoplasmic Ca^{2+} (ref. 4), which can activate cell responses without cell death^{5,6}. Here we report the use of a monoclonal antibody to the terminal complement component C9 (refs 7, 8, 13), quantified by ¹²⁵I and visualized by fluorescein, to demonstrate a protection mechanism in polymorphonuclear leukocytes (PMNs) attacked by complement, involving removal of the attack complex by vesiculation. Concomitantly, there is a Ca^{2+} -dependent activation of reactive oxygen metabolite production without cell lysis. These findings have important implications in the evolutionary and pathological significance of the terminal components of the complement pathway.

Two experimental approaches were adopted to establish the existence of a protection mechanism in PMNs attacked by complement. First, the fate of the membrane attack complex (C5b6789_n) was investigated using ¹²⁵I-labelled or fluorescein-labelled monoclonal antibody against C9 (Figs 1-3). Second, the ability of the membrane attack complex to activate cells in the absence of cell lysis was established by observing its effect on reactive oxygen metabolite production in PMNs (Fig. 4), a response known to be activated by an increase in intracellular Ca^{2+} (refs 9 and 26). It was therefore necessary first to form stable cell intermediates bearing sub-lytic amounts of the attack complex (sub-lytic PAC1-9).

Polymorphonuclear leukocytes, sensitized with a rabbit anti-PMN antibody, were first incubated for 10 min at 37 °C with a human serum depleted specifically of C9 (NHS-C9) using a solid-phase monoclonal antibody to C9 (ref. 8) at a dilution which, in the presence of C9, caused <10% end-point lysis (detected using trypan blue exclusion and lactate dehydrogenase release). These PAC1-8 intermediates were then incubated at 0 °C with C9 (10 μ g l⁻¹; 140 nM). The PAC1-9 intermediates

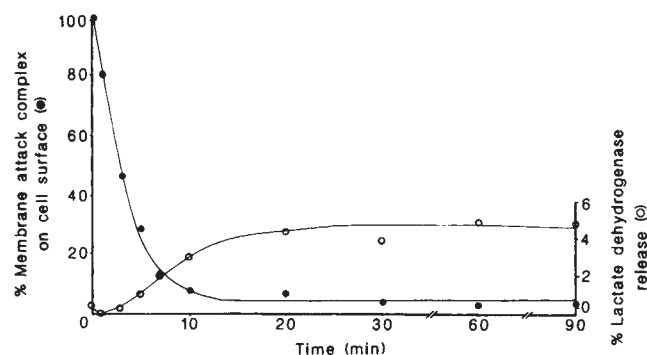


Fig. 1 Removal of the membrane attack complex from polymorphonuclear leukocytes. Rat PMNs sensitized with antibody were incubated for 15 min at 37°C with human serum depleted specifically of C9 at a dilution (1:25 in Krebs HEPES, pH 7.4) which in the presence of C9 caused ~10% end-point lysis. The PAC1-8 intermediates thus formed were washed, resuspended to a cell concentration of 10^7 per ml and incubated with pure C9 ($10 \mu\text{g ml}^{-1}$, 140 nM) at 0°C for 10 min. The sub-lytic PAC1-9 intermediates thus formed were washed and stored at 0°C. Sub-lytic PAC1-9 (10^7 cells ml^{-1}) were placed at 37°C and 250- μl aliquots (2.5×10^6 cells) removed at defined time intervals and immediately placed on ice. Radiolabelled monoclonal antibody against C9 (MC47, 5×10^4 d.p.m. ^{125}I) was added and the cells incubated at 0°C for a further 10 min. The radioactivity in the washed cell pellet was then measured. Results are expressed as percentage of radiolabelled antibody bound at time 0 (no incubation at 37°C) and are the means of two observations. Cell death was estimated by measuring lactate dehydrogenase release and by trypan blue exclusion.

thus formed were stable at 0°C, and end-point lysis on incubation at 37°C was <5%. ^{125}I -labelled monoclonal antibody against C9 (coded MC47), which had previously been shown to bind to C9 after its insertion into the cell membrane attack complex¹⁰, showed that >2,000 MC47 antibody molecules per cell were bound specifically in the sub-lytic PAC1-9 intermediates. This indicated that many attack complexes were required to cause cell lysis, suggesting that not all the attack complexes on the cell surface were functional lesions^{11,12}. Nonspecific binding of anti-C9 antibody MC47 to PMNs alone or to PAC1-8 intermediates was less than 5% of its binding to PAC1-9. Preincubation of PMNs with C9 alone caused no detectable increase in the binding of antibody ^{125}I -MC47. The stability of the sub-lytic PAC1-9 intermediates at 0°C was confirmed by the fact that no decrease in ^{125}I -MC47 binding was detected after the intermediates had been stored at 0°C for 60 min. Similarly, using fluorescein-labelled anti-C9 antibody MC47, >95% of the sub-lytic PAC1-9 intermediates had visible fluorescence, and thus membrane attack complexes, on their surfaces.

The fate of the attack complex on incubation of sub-lytic PAC1-9 at 37°C was quantified using ^{125}I -MC47 (Fig. 1) and visualized using fluorescein MC47 antibody (Fig. 2). Binding of ^{125}I -MC47 antibody to sub-lytic PAC1-9 intermediates decreased rapidly at 37°C, being reduced to 50% within 3.5 min and to <10% of the original binding within 10 min (Fig. 1). Removal of the attack complexes containing C9 at 37°C was confirmed by visualizing it with fluoresceinated MC47 (Fig. 2). Before incubation at 37°C, fluorescence was uniformly distributed over the cell surface (Fig. 2a). Within 2 min at 37°C the cell-surface MC47 fluorescence coalesced into bright patches (Fig. 2c). This was followed by the formation of highly fluorescent vesicles (Fig. 2e, g) which were shed into the surrounding medium. Internalization of fluorescent MC47 vesicles was also observed in many cells (Fig. 2c). These events were not the result of capping induced by the anti-C9 monoclonal antibody itself, as patching and vesiculation of cell membrane fluorescence were also observed in sub-lytic PAC1-9 intermediates where C9 had been labelled directly with fluorescein. Use of labelled antibody to visualize C9 provides a larger fluorescence signal, making it easier to record the protection mechanism.

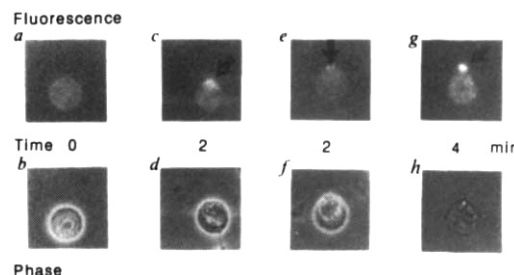


Fig. 2 Visualization of removal of membrane attack complexes from polymorphonuclear leukocytes. a, c, e, g, Fluorescence; b, d, f, h, phase contrast. Before incubation at 37°C, sub-lytic PAC1-9 intermediates exhibited a diffuse cell-surface fluorescence (a). Within 2 min of being placed at 37°C, patching of fluorescence on the cell surface and the appearance of fluorescent vesicles within the cell were observed (c). At 4 min, highly fluorescent vesicles were observed on many cells, which were subsequently released into the surrounding medium (e, g). After incubation at 37°C for 60 min, <5% of cells had detectable surface fluorescence (h). PAC1-8 intermediates similarly incubated with fluorescent antibody had no detectable cell-surface fluorescence.

Methods. Rat PMNs bearing sub-lytic amounts of the membrane attack complex (sub-lytic PAC1-9) were prepared as described in Fig. 1 legend. Sub-lytic PAC1-9 (10^7 per ml) were then incubated with fluorescein-conjugated antibody against C9 (MC47, $20 \mu\text{g ml}^{-1}$) at 0°C for 10 min. The cells were washed at 0°C, then incubated at 37°C. Portions were removed at intervals, placed at 0°C and cell-surface fluorescence was examined using a Zeiss photomicroscope.

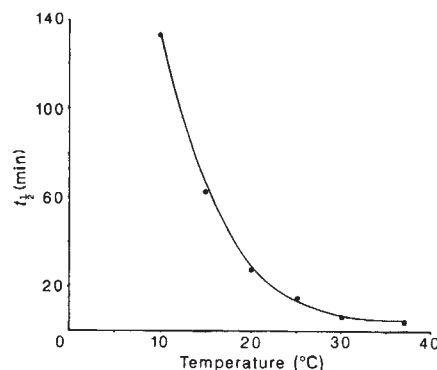


Fig. 3 Temperature dependence of removal of membrane attack complexes. Rat PMNs bearing sub-lytic amounts of attack complex (sub-lytic PAC1-9) were incubated at various temperatures in the range 10–37°C, and binding of radiolabelled antibody against C9 was measured at intervals as described in Fig. 1 legend. The decrease in binding of radiolabelled antibody appeared to be exponential down to 10% of original binding. The time taken for binding of radiolabelled antibody to decrease to 50% of original ($t_{1/2}$) was estimated from duplicate measurements and plotted against incubation temperature.

The removal of attack complexes from the cell membrane of sub-lytic PAC1-9 intermediates was markedly temperature dependent (Fig. 3). At 10°C the time for loss of 50% ^{125}I -MC47 binding was 133 min, compared with 3.5 min at 37°C. At 0°C there was no detectable loss of antibody MC47 binding over 3 h. C9 was required to remove the attack complexes, as removal of the C5-8 complexes without C9 was slow, 50% being lost within 42 min at 37°C; this was demonstrated by measuring binding of ^{125}I -C9 to PAC1-8 intermediates.

We have shown previously that formation of the attack complexes on PMNs causes a stimulation of reactive oxygen metabolite production (O_2^- , $\cdot\text{OH}$, H_2O_2 , OCl^- , $^1\text{O}_2$), detectable within 20 s using luminol-dependent chemiluminescence^{5,9,14}. This activation requires C9 and extracellular Ca^{2+} , occurring many minutes before significant cell lysis is detectable. Addition of C9 to PAC1-8 intermediates at 37°C, thereby forming sub-lytic PAC1-9 intermediates, also caused a rapid, transient increase in reactive oxygen metabolite production (Fig. 4a). In contrast,

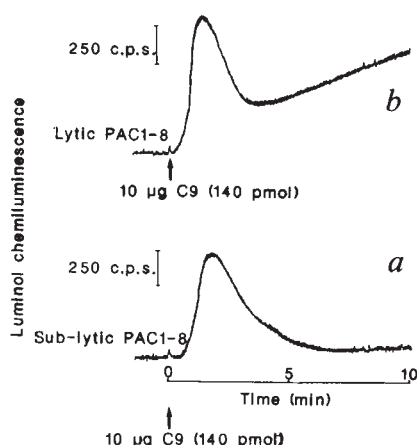


Fig. 4 Stimulation of the production of polymorphonuclear leukocyte reactive oxygen metabolites by sub-lytic amounts of the membrane attack complex. *a*, Sub-lytic PAC1-8 + C9; *b*, lytic PAC1-8 + C9. End-point lysis was calculated by measurement of lactate dehydrogenase release after 90 min incubation at 37 °C and was 4.2% for *a* and 39.5% for *b*.

Methods. PAC1-8 intermediates were produced by incubation of antibody-sensitized PMNs with C9-depleted serum at a dilution which, in the presence of C9, caused <10% end-point lysis (sub-lytic PAC1-8) or ~40% end-point lysis (lytic PAC1-8). Intermediates (5×10^6 in 0.5 ml Krebs HEPES containing 11 µM luminol) were then incubated at 37 °C in the reaction chamber of a computerized chemiluminometer¹⁴. C9 (10 µg in 0.5 ml medium at 37 °C) was then added and luminescence monitored.

addition of C9 to PAC1-8 intermediates to form PAC1-9 lytic intermediates (end-point lysis ~40%) resulted in a biphasic luminol-dependent chemiluminescence (Fig. 4*b*) similar to that previously observed using whole serum at lytic doses of attack complex⁵. The first phase is thus a sub-lytic event, ceasing on the removal of the attack complexes from the cell surface, whereas the second phase is a pre-lytic or lytic event and may be the result of myeloperoxidase release^{14,15}. The rapid first phase parallels the rapid rise in intracellular free Ca^{2+} induced by C9 incorporation into the attack complex^{4,9,26}. Thus, this increase in intracellular Ca^{2+} may activate not only metabolic pathways such as that responsible for reactive oxygen metabolite production, but also membrane recovery processes preventing cell death in nucleated cells at sub-lytic levels of complement attack⁶.

Although the pathways for complement activation are well known, the precise mechanism by which the membrane attack complex affects nucleated cells remains controversial^{6,16-18}. The results presented here shed new light on complement action on nucleated cells. Cell death caused by complement belongs to a class of phenomena defined as 'threshold'¹⁹, where cell death is dependent on a critical chemical or osmotic threshold within the cell or cell membrane. This threshold may not be reached for several minutes after formation of the attack complex on the membrane of an individual cell²⁰, yet chemical events can occur within the cell in seconds as a result of changes in the intracellular concentration of ions such as free Ca^{2+} (refs 5, 6, 19, 26). In cells which can protect themselves before the threshold is reached (Figs 1, 2), chemical and morphological changes¹⁰ can still occur and yet the cell can survive. Consequently, any inhibitor of the recovery process will increase the number of cells passing through the threshold and thus increase cell lysis. We have recently shown that adenosine, an activator of adenylyl cyclase in PMNs, inhibits removal of the attack complex and concomitantly increases cell lysis²¹.

Vesiculation and activation of cells by complement have been reported previously^{5,22,23}, but such studies could not distinguish whether these events were sub-lytic or pre-lytic. Our findings of a mechanism for removal of membrane attack complexes from the cell surface, together with activation of intracellular events without cell lysis, provide a new basis for understanding the

possible role of complement in the pathogenesis of immune-based disease^{6,18,24}. The recovery process in PMNs is quite distinct from resistance of prokaryotes to complement attack, the latter being due to failure of the attack complex to insert into the outer or inner membrane of the bacterium²⁵. The establishment of a recovery process, via removal of the attack complex, in nucleated cells other than PMNs will provide a new perspective in the understanding of a controversial problem—the natural function and evolutionary significance of C9, the terminal component of complement.

We thank the MRC and the Arthritis and Rheumatism Council for financial support, also Drs M. B. Hallett and J. P. Luzio for helpful discussions, and the latter and Dr K. Siddle for help with the production of C9 monoclonal antibodies.

Received 17 January; accepted 24 July 1985.

- Boyle, M. D. P., Ohanian, S. H. & Borsos, T. J. *J. Immun.* **116**, 1276-1279 (1976).
- Ohanian, S. H. & Schlager, S. I. *CRC crit. Rev. Immun.* **1**, 165-209 (1981).
- Sissons, J. G. & Oldstone, M. B. A. *Adv. Immun.* **29**, 209-260 (1980).
- Campbell, A. K., Daw, R. A. & Luzio, J. P. *FEBS Lett.* **107**, 55-60 (1979).
- Hallett, M. B., Luzio, J. P. & Campbell, A. K. *Immunology* **44**, 569-577 (1981).
- Campbell, A. K. & Luzio, J. P. *Experientia* **37**, 1110-1112 (1981).
- Morgan, B. P., Daw, R. A., Siddle, K., Luzio, J. P. & Campbell, A. K. *J. immun. Meth.* **84**, 269-281 (1983).
- Morgan, B. P., Campbell, A. K., Luzio, J. P. & Siddle, K. *Clin. chim. Acta* **134**, 85-94 (1983).
- Hallett, M. B. & Campbell, A. K. *Nature* **259**, 155-157 (1982).
- Morgan, B. P., Sewry, C. A., Siddle, K., Luzio, J. P. & Campbell, A. K. *Immunology* **52**, 181-188 (1984).
- Koski, C. L., Ramm, L. E., Hammer, C. H., Mayer, M. M. & Shin, M. L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3816-3820 (1983).
- Ramm, L. E., Whitlow, M. B., Koski, C. L., Shin, M. L. & Mayer, M. M. *J. Immun.* **131**, 1411-1415 (1983).
- Morgan, B. P., Luzio, J. P. & Campbell, A. K. *Biochem. biophys. Res. Commun.* **118**, 616-622 (1984).
- Campbell, A. K., Holt, M. & Patel, A. *Recent Adv. clin. Biochem.* **3**, 1-30 (1985).
- Hallett, M. B. & Campbell, A. K. *Biochem. J.* **216**, 459-465 (1983).
- Biesecker, G. *Lab. Invest.* **49**, 237-250 (1983).
- Esser, A. F. in *Biological Membranes* Vol. 4 (ed. Chapman, D.) 276-325 (Academic, New York, 1982).
- Lachmann, P. J. *Nature* **305**, 473-474 (1983).
- Campbell, A. K. *Intracellular Calcium: Its Universal Role as Regulator* (Wiley, Chichester, 1983).
- Edwards, S. W., Morgan, B. P., Hoy, T. G., Luzio, J. P. & Campbell, A. K. *Biochem. J.* **206**, 195-202 (1983).
- Roberts, P. A., Morgan, B. P. & Campbell, A. K. *Biochem. biophys. Res. Commun.* **126**, 692-697 (1985).
- Richardson, P. J. & Luzio, J. P. *Biochem. J.* **186**, 897-906 (1980).
- Imagawa, D. K., Osifchin, N. E., Paznekas, W. A., Shin, M. L. & Mayer, M. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6647-6651 (1983).
- Morgan, B. P., Campbell, A. K. & Compston, A. *Lancet* **ii**, 251-255 (1984).
- Joiner, K. *et al. Fedn Proc.* **44**, 552 (1985).
- Campbell, A. K. & Hallett, M. B. *J. Physiol., Lond.* **338**, 537-550 (1983).

Sequence polymorphism of HLA *DRB1* alleles relating to T-cell-recognized determinants

J. S. Cairns, J. M. Curtsinger, C. A. Dahl, S. Freeman, B. J. Alter & F. H. Bach

Immunobiology Research Center, Departments of Laboratory Medicine/Pathology and Surgery, Box 724, Mayo Building, University of Minnesota, Minneapolis, Minnesota 55455, USA

HLA class II molecules are a highly polymorphic family of dimeric cell-surface proteins primarily involved in regulating T-cell responses to extrinsic antigens. To define regions of class II molecules involved in T-cell recognition, we have now compared sequences of three HLA *DRB* cDNA clones obtained from cells that all express the same serologically defined determinants but differ in terms of T-cell-recognized specificities. The comparisons indicate that very few (one to four) nucleotides differ between what are almost certainly alleles of the *DRB1* locus. All differences were in the first domain of the molecule and all localized to a region from amino acids 71-86. Because all differences were found only in this region of the molecule, and because DR α -chains seem to be relatively non-polymorphic¹, these positions in the DR β -chain must have a major role in influencing T-cell recognition of the DR molecule.

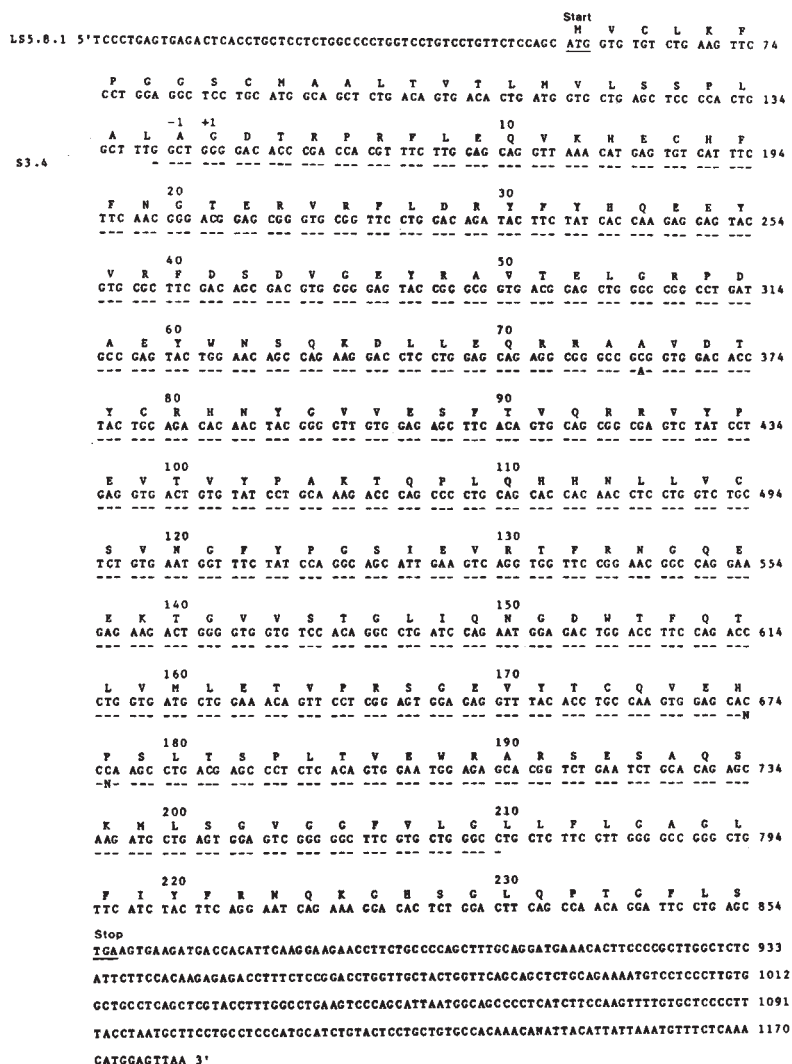


Fig. 1 Sequences for clones LS5.8.1 and S3.4. Both LS5.8.1 and S3.4 were initially isolated from cDNA libraries constructed from membrane-bound (LS5.8.1) or cytoplasmic (S3.4) RNA. cDNA was synthesized using oligo(dT) to prime the first strand, and nascent mRNA nicked with RNase H to prime the second strand. cDNA was 'blunt-ended' with DNA polymerase and tailed with dCTP using terminal deoxynucleotidyltransferase. cDNA was annealed with *Pst*I-cut, dG-tailed pUC9, and introduced into *Escherichia coli* strain JM83. Each library, which contained ~7,000 colonies, was screened with a 790-base pair *Sst*I/*Hind*III fragment of a *DRβ* cDNA clone isolated by Long *et al.*⁷ This probe contains almost the entire coding region from this clone and in the conditions of hybridization used, reacts most strongly with *DRβ* cDNA but also reacts weakly with *DPβ* and *DQβ*. The inserts from these clones, or fragments thereof, were subcloned into the single-stranded vector M13mp19. Subclones were either sequenced directly or truncated subclones were generated by the method of Dale *et al.*¹⁷. Sequencing was done using the dideoxy chain termination method¹⁸. N in the S3.4 sequence represents an ambiguous area of the sequencing gel. In addition, ~90 nucleotides at the 5' end of this insert have not been sequenced. Where the sequence of S3.4 is identical to that of LS5.8.1, a dash is shown below the LS5.8.1 sequence.

Serological and molecular cloning experiments have shown that there are three families of class II molecules within the HLA complex, termed DP, DQ and DR. For the *DR4-DRw53* haplotype which we have studied, within the *DR* family, one α and three complete β genes are known; one β gene is a pseudogene². At least two *DR αβ* dimers are expressed. The serological determinant DR4 is associated with the $\alpha\beta$ 1 dimer whereas the *DRw53* determinant is associated with the $\alpha\beta$ 2 dimer. To date, comparisons of *DRβ* nucleic acid sequences have been made between cells differing both serologically and in terms of T-lymphocyte-recognized determinants, making it impossible to ascribe determinants recognized by T lymphocytes to particular polymorphic sequences. To minimize differences to those potentially involved in T-cell recognition, we have compared sequences of what appear to be *DRβ*1 cDNA clones from cells which differ with respect to determinants, designated as 'Dw', recognized by allogeneic T cells, but which share the serologically defined specificity DR4.

Among cells which express DR4, at least five distinct Dw subtypes have been defined, designated Dw4, Dw10, Dw13, Dw14 and Dw15 (ref. 3). Clones LS5.8.1 and S3.4 were obtained from cDNA libraries of two homozygous typing cells, LS40 and SSTO, expressing Dw14 and Dw13, respectively. The sequences of these clones are shown in Fig. 1. The clones differ by a single nucleotide, resulting in an alanine in LS5.8.1 (Dw14) and a glutamic acid in S3.4 (Dw13) at amino-acid position 74 (Table 1, lines a, b).

LS5.8.1 differs by only three nucleotides (two amino acids, Table 1, lines a, c) from an unpublished *DRβ* sequence

(E. Long, B. Mach *et al.*, personal communication) obtained from a *DR4*⁺ cell, referred to as 'DR4.6'. To make this sequence comparison more meaningful, we used cloned cytotoxic T lymphocytes that differentiate between the DR products (probably *DRβ*1)⁴ of Dw4 and Dw14, and established that DR components of Dw4 are associated with the *DR4* haplotype of DR4.6. As shown in Table 2, both LS40 and DR4.6 served as effective targets for cloned cytotoxic T lymphocytes which recognize both Dw4- and Dw14-associated determinants of DR4.6 and LD40; only DR4.6 was lysed by cloned cytotoxic T lymphocytes recognizing Dw4 but not Dw14 determinants.

Based on the high degree of sequence homology between the *DRβ* clones from Dw14, Dw13 and Dw4 cells described, it is

Table 1 Summary of amino-acid differences between *DRβ*1 alleles

Clone	Cell	DR	Dw	Amino-acid position		
				71	74	86
a LS5.8.1	LS40	β 1	14	Arg	Ala	Val
b S3.4	SSTO	β 1	13	Arg	Glu	Val
c	DR4.6*	β 1	4	Lys	Ala	Gly
d	Preiss†	β 1	4	Lys	Ala	Gly
e	Preiss†	β 2	4	Arg	Glu	Val

* Taken from the sequence of a *DRβ* cDNA clone obtained by B. Grubenmann, E. O. Long and B. Mach, in preparation.

† Taken from ref. 5.

Table 2 Lysis of DR4,6, LS40 and other targets by cytotoxic T-lymphocyte clones

Clone*	Target				
	DR4/6	LS40	BO (Dw4 homozygous typing cell)	K.Blo (Dw3,4)	J.Nor (Dw2,3)
KD15	43.0 ± 4.4†	0.8 ± 1.4	35.7 ± 4.6	40.5 ± 6.3	2.1 ± 1.3
KD33	59.2 ± 6.6	2.1 ± 1.2	27.2 ± 3.9	40.0 ± 2.8	10.1 ± 2.0
KD48	55.4 ± 8.3	56.3 ± 9.5	16.1 ± 2.7	23.4 ± 2.7	1.1 ± 2.5

The cell-mediated lympholysis (CML) assay was performed as follows. Target cells were labelled with 0.25 mCi chromium-51 for 1 h, washed three times with cold RPMI-1640 containing 20% pooled human serum, and adjusted to 1×10^4 cells ml^{-1} . The cloned effectors were resuspended in interleukin-2-free culture medium the day before testing. Effectors were adjusted to the appropriate concentration and 100 μl effectors + 100 μl targets were added to V-bottom microtitre plates, spun at 500 r.p.m. for 5 min and incubated in a 37 °C, 5% CO_2 humidified environment for 4 h. The plates were then spun at 1,000 r.p.m. for 10 min and 150 μl of supernatant was aspirated and placed in scintillation vials. Ready-Solv HP (Beckman 566436) (2.5 ml) was added to each vial and the samples were counted in a β scintillation counter. Spontaneous release of ^{51}Cr was assessed by incubating target cells without effector cells and maximum release of the isotope was assessed by incubating target cells with 0.1% hexadecyltrimethylammonium bromide. Results were calculated as follows:

$$\% \text{ Cytotoxicity} = \left(\frac{\text{c.p.m. experimental wells} - \text{c.p.m. spontaneous release}}{\text{c.p.m. maximum release} - \text{c.p.m. spontaneous release}} \right) \times 100$$

* KD15 and KD33 are Dw4 specific; KD48 lyses Dw4 and Dw14 targets⁴.

† % CML $\pm$ s.d. The effector: target ratio is 20:1.

extremely likely that these clones represent transcripts of alleles of a single locus. The clones presumably represent transcripts of the *DRβ1* locus for the following reasons. Speis *et al.*⁵ have compared sequences of the first domains of known (based on amino-acid sequencing of characterized proteins) *DRβ1* and *DRβ2* genomic clones from Priess (a *DR4-Dw4* homozygous typing cell) and have found 19 amino-acid differences between them. The *DR4-Dw4 β1* sequence from Priess is identical to the sequence of *DR4,6*. Our allelic sequences, because they differ by 2–3 amino acids from the *DRβ1* sequence of Priess, but by 16–17 amino acids from the *DRβ2* sequence, probably represent *DRβ1* sequences. Also, at the positions where the *DRβ1* sequences of Dw14, Dw13 and Dw4 differ from each other, one of the alternative amino acids and the corresponding codon are also present at the same position in the *DRβ2* sequence. Whether this has occurred as the result of accumulated mutation in the *DRβ1* gene, or through a gene conversion-like mechanism, is unclear. If gene conversion-like events have occurred, it is thus possible that the *DRβ2* locus functioned as a donor gene responsible for the substitutions seen between *DRβ1* alleles; however, a *DRβ* pseudogene sequence reported by Larhammar *et al.*² also contains codons identical to those of *DRβ2* at these positions and could therefore also have functioned as a donor gene.

The conclusion that these sequences represent *DRβ1* transcripts is also consistent with the observation that there is an approximately 10-fold higher level of *DRβ1* than *DRβ2* expressed on the surface of B cells transformed by Epstein-Barr virus^{5,6}; that finding may well be reflected in levels in the messenger RNA pool. We have now sequenced all or part of three different *DRβ* cDNA clones from our LS40 library and, although there is evidence that transcripts of this gene are differentially processed (J.S.C., C.A.D., J.M.C. and F.H.B., in preparation), all are apparently transcripts of a single locus. The clone sequenced by Long *et al.* is also a member of a highly represented family of *DRβ* clones (group 2 in ref. 7).

As it is likely that these sequences represent transcripts of the *DRβ1* locus, these data suggest that a few amino-acid differences may profoundly affect how the molecule is recognized by T lymphocytes. This situation is not without precedent; relatively minor amino-acid substitutions in the heavy chain of the HLA-A2 molecule^{8,9} and in mutants of the K^b molecule¹⁰ drastically alter T-cell recognition of those molecules.

Structural models of the DR molecule have placed residues 71, 74 and 86 on the outer face of the DR molecule¹¹. Residues 71 and 74 are included in the third hypervariable region of the DR β -chain, the only proposed α -helical region in the first domain¹¹. Amino-acid substitutions in the I-A mutant bm12 are

in this same region¹². If determinants of the *DRαβ1* dimer are involved in restricted recognition of foreign (nominal) antigens, and there is some evidence for this^{13,14}, these same amino-acid differences among *DRβ1* must be involved.

Whether the changes in DNA sequence noted here are directly responsible for encoding T-cell-recognized determinants or whether they are responsible for conformational changes involving other parts of the molecule which are recognized by T cells remains unclear. We would also expect to find other sequence differences that could result in determinants recognized by T cells. We postulate, however, as we have discussed recently^{15,16}, that the allelic differences reported here in a naturally evolving population may be of primary importance evolutionarily and functionally for T-lymphocyte recognition.

We thank Drs J. Strominger and E. Long for making unpublished data available and Dr E. Long, Professor B. Mach and Professor P. Peterson for their donation of *DRβ* cDNA clones. J.S.C. thanks D. Larhammar, L. Rask and B. Laurent and Professor Peterson and B.J.A. thanks Dr Tomi Meo for assistance during initial stages of this investigation. We thank the University of Minnesota Clinical MLC Laboratory for growing and typing the lymphoblastoid cell lines used in this study, and Ms N. Andresen for typing this manuscript. This is paper 421 from the Immunobiology Research Center, University of Minnesota, Minneapolis, Minnesota 55455. Supported by the Harry Kay Charitable Foundation, NIH grants AI 17697, AI 183326, AI 19007 and contract ONR N14-85-K-0004.

Received 29 May; accepted 9 July 1985.

1. Crumpton, M. J. *et al.* in *Histocompatibility Testing 1984* (ed. Albert, E.) 29–37 (Munksgaard, Copenhagen, 1984).
2. Larhammar, D., Servenius, B., Rask, L. & Peterson, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1475–1479 (1985).
3. Reinsmoen, N. L. & Bach, F. H. *Hum. Immun.* **4**, 249–258 (1982).
4. Reinsmoen, N. L. & Bach, F. H. *Hum. Immun.* (in the press).
5. Spies, T., Sorrentino, R., Boss, J. M., Okada, K. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
6. Yabe, T., Suzuki, M., Satake, M., Juji, T. & Hamaguchi, H. *Immunogenetics* **20**, 155–167 (1984).
7. Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389–394 (1983).
8. Biddison, W. E., Ward, F. E., Shearer, G. M. & Shaw, S. J. *Immun.* **124**, 548–552 (1980).
9. Goulmy, E., van Leeuwen, A., Blokland, F., van Rood, J. J. & Biddison, W. E. *J. exp. Med.* **155**, 1567–1572 (1982).
10. Nairn, R., Yamaga, K. & Nathanson, S. G. *A. Rev. Genet.* **14**, 241–277 (1980).
11. Norcross, M. A. & Kanehisa, M. *Scand. J. Immun.* (in the press).
12. McIntyre, K. R. & Seidman, J. G. *Nature* **308**, 551–553 (1984).
13. Reinsmoen, N. L., Volkman, D. J., Bach, F. H. & Fauci, A. S. *Fedn Proc.* **43**, 1819 (1984).
14. Qvistad, E., Thorsby, E., Reinsmoen, N. L. & Bach, F. H. *Immunogenetics* **20**, 583–588 (1984).
15. Nicklas, J. N., Noreen, H. N., Segall, M. & Bach, F. H. *Hum. Immun.* **13**, 95 (1985).
16. Bach, F. H. *Immun. Today* **6**, 89–94 (1985).
17. Dale, R., McClure, B. & Houchins, J. P. *Plasmod* **13**, 13–40 (1985).
18. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463 (1977).

Simian virus 40 replication in adenovirus-transformed human cells antagonizes gene expression

Jane S. Lebkowski, Suzanne Ciancy & Michele P. Calos

Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

Simian virus 40 (SV40) replicates efficiently in monkey kidney cells. However, we have now found that SV40-based vectors transfected into most human cells replicate poorly, if at all. In contrast, strong SV40 replication is observed in human embryonic kidney (HEK) cells transformed with the adenovirus early region, but not in untransformed HEK cells. Vector replication in adenovirus-transformed cells is dependent on the presence of the SV40 origin of replication and large-T antigen. However, vigorous replication occurs at levels of large-T antigen that are undetectable by immunofluorescence. These data suggest that the adenovirus oncogenes create a replication-permissive environment to which the SV40 replicon responds. Furthermore, replication and gene expression seem to be antagonistic on our vectors. High levels of large-T antigen are observed only when vector replication is blocked by mutations in the gene for large-T antigen or the origin of replication, or by direct inhibition of DNA polymerase with aphidicolin.

We have been examining mutagenesis in mammalian cells using a shuttle vector approach^{1,2}. Extension of these experiments to human cells requires efficient replication of a shuttle vector in the human cell environment. As human cells are considered semi-permissive for SV40³, we examined replication of an SV40-based vector in various human cell lines. The vector, pJYMib, contains the entire SV40 genome inserted into the *Bam*HI site of the pBR322 derivative pML (ref. 4) and also contains the *lacI* gene². Vector DNA was introduced into human cells in culture using the calcium phosphate co-precipitation technique⁵. After replication for 1–5 days, plasmid DNA was collected by the Hirt procedure⁶ and digested with *Mbo*I to distinguish replicated from input DNA². The *Mbo*I digest was run on an agarose gel, blotted to a filter and probed with ³²P-labelled pJYMib DNA. The intensity of the *Mbo*I pattern gives a direct measure of the amount of vector replication in the mammalian cell.

This replication assay was used on various human cells, including SV40-transformed xeroderma pigmentosum skin fibroblasts (Human Genetic Mutant Cell Repository NIGMS GM 4312A), 143 thymidine kinase-negative osteosarcoma cells⁷, HeLa cervical carcinoma cells and primary HEK cells derived from 9–12-week-old human fetuses⁸ (kindly provided by E. Major), but none showed any indication of pJYMib replication except HeLa, where a low level of replication was detectable (see below).

Because large-T antigen is required for SV40 replication, it was possible that poor expression of SV40 large-T antigen might be responsible for the observed block to SV40 replication. It has been reported^{9–12} that transfected DNA is strongly expressed in the 293 cell line¹³, an adenovirus-transformed derivative of HEK. The effect is apparently mediated by the adenovirus *E1a* gene products, which are efficiently expressed in 293. It therefore seemed possible that SV40 large-T antigen might be expressed at a high level in 293. Accordingly, we transfected pJYMib into late-passage 293 cells (from A. Berk via C. Hsu) and assayed replication. Figure 1 shows pJYMib indeed replicated very strongly in these cells². Its replication in 293 was much stronger than that detected in CV-1, the African green monkey kidney line commonly used for propagation of SV40. Replication levels in 293 were comparable to those observed in COS-7, an SV40-transformed CV-1 derivative which produces large-T antigen constitutively¹⁴. Replication increased over at least 72 h. The

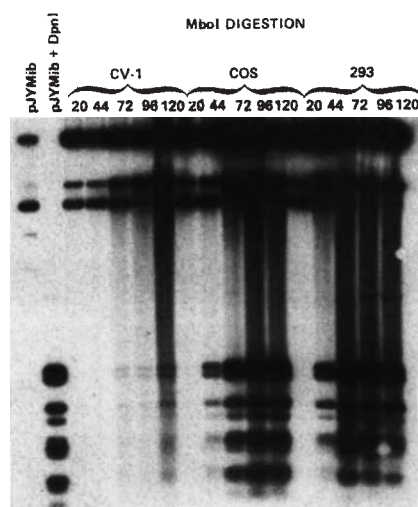


Fig. 1 Time course of pJYMib replication. 60-mm dishes of CV-1, COS-7 and 293 cells were transfected with 2 µg of pJYMib DNA using the calcium phosphate technique⁵. The cells were collected 20, 44, 72, 96 and 120 h after transfection. Plasmid DNA was purified by Hirt extraction⁶ and one-third of the sample was digested with *Mbo*I (New England Biolabs) and loaded on a 1% agarose gel. Uncut pJYMib DNA (5 ng) was loaded at the left as a marker for input DNA. pJYMib (25 ng) digested with *Dpn*I (Boehringer Mannheim) was used as the marker for the *Mbo*I digestion pattern. Bacterially grown DNA is sensitive to *Dpn*I digestion, but only DNA which has been replicated in the mammalian cells and lost its bacterial adenine methylation is sensitive to *Mbo*I. The bands near the top of the gel represent residual input plasmid DNA, while the bands at the bottom reflect DNA replicated in the mammalian cell.

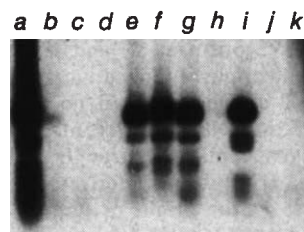


Fig. 2 Comparative vector replication. 60-mm dishes of cells were transfected and DNA collected at 72 h and digested with *Mbo*I as in Fig. 1. Only the *Mbo*I patterns are shown. Lanes a–f represent pJYMib DNA transfected into COS-7 (a), CV-1 (b), HEK (c), HeLa (d), 293 (e) and JW-2 (f) cells. Lanes g–k represent various vectors transfected into 293 cells. g, pJYMib; h, pJYMib, plus 20 µM aphidicolin added 20 h post-transfection; i, pMK16/ori-; j, pMK16/ori-; k, pMK16/SV80.

high yield of replicated SV40 DNA in 293 was in contrast with all the other human cell lines tested and occurs even though the transfection efficiencies of the cell lines are comparable (see below).

293 was created by transfection of HEK cells with sheared fragments of adenovirus-5, resulting in a transformant which contains and expresses the early genes *E1a* and *E1b*, the transforming functions of adenovirus¹³. Figure 2 (lanes a–e) shows the relative replication levels of pJYMib at 72 h in COS-7, CV-1, HEK, HeLa and 293. Replication in HeLa was markedly below that in 293, and replication in HEK was undetectable. As HEK cells are derived from the same tissue type as 293, the adenovirus early genes may be responsible for the replication permissivity of 293. Further evidence implicating the adenovirus oncogenes comes from our study of the JW-2 line (from P. Gallimore,

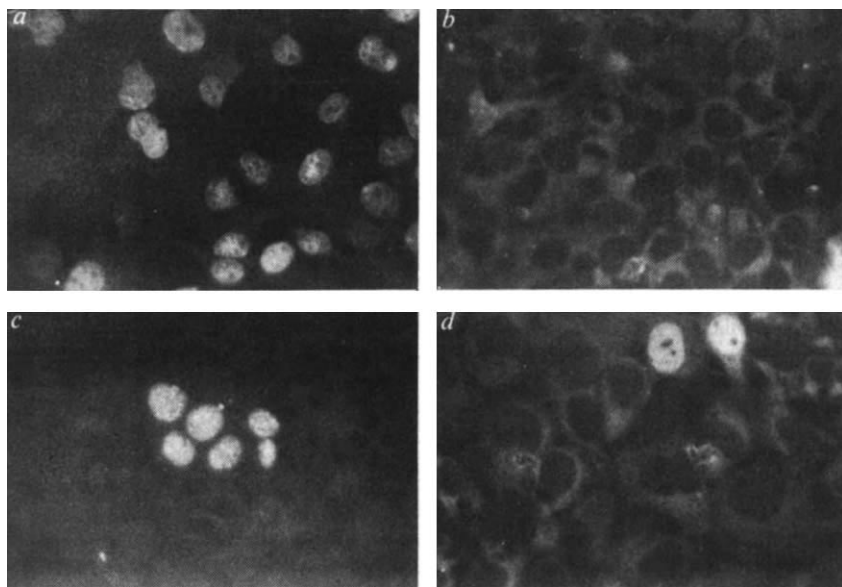


Fig. 3 Large-T antigen immunofluorescence. *a*, COS-7; *b*, 293 transfected with pMK16/SV40; *c*, 293 transfected with pMK16/ori⁻; *d*, 293 transfected with pMK16/SV40, plus aphidicolin (see text).

Methods. 60-mm dishes of cells were fixed with methanol (Mallinkrodt) for 20 min 48–72 h after transfection. 10 μ l of 1:16 dilution of hamster anti-SV40 tumour serum (gift of M. Botchan) in phosphate-buffered saline (PBS) was applied to a 1-cm diameter section of the dish and incubated in a CO₂ incubator for 30 min. After two washes with PBS, 10 μ l of a 1:10 dilution of fluorescein isothiocyanate-labelled goat anti-hamster IgG (Cappel Laboratories) was applied to the circle and incubated at 37°C for a further 30 min. After two washes with PBS, the cells were examined for indirect immunofluorescence with a Zeiss standard epifluorescence microscope ($\times 10$ ocular) and a $\times 40$ Panneofluor objective. Cells were photographed on Kodak Tri-X film.

courtesy of J. Williams). JW-2 is an independently derived HEK transformant constructed by microinjection of the *Eco*RI-C restriction fragment containing the leftmost 16.5% of adenovirus-12 into HEK cells¹⁵. Replication levels in JW-2 are comparable to those in 293 (Fig. 2, lane *f*). Therefore, two independently derived adenovirus-transformed HEK lines both support high levels of SV40 replication. This result, together with the failure of all the other human lines to replicate SV40 efficiently, argues for a role for the adenovirus gene products.

We performed genetic experiments to verify that large-T antigen and the SV40 origin were required for SV40 replication in 293. A plasmid, pSVi ori⁻, lacking all SV40 sequences², was assayed for replication and produced a negative result. Similarly, pSVi2 (ref. 1), pSV2gptib (ref. 2) and several other plasmids containing the SV40 origin of replication, but lacking the coding region for large-T antigen, showed no replication. Plasmids pJYMib and pMK16/SV40¹⁶, which contain both the viral origin and large-T antigen, replicated strongly and at similar levels (Fig. 2, lanes *g*, *i*). However, derivatives of pMK16/SV40 possessing defects in the origin or large-T antigen failed to replicate. Plasmid pMK16/ori⁻, which contains a mutation at the *Bgl*I site of SV40 which inactivates the origin¹⁶, showed no replication in our assay (Fig. 2, lane *j*). Similarly, pMK16/SV80, which contains a point mutation affecting amino-acid 147 in the DNA binding region of large-T antigen (W. Gish and M. Botchan, personal communication), also failed to replicate in 293 (Fig. 2, lane *k*). These results suggest that both the origin and large-T antigen are required for SV40 replication in 293 cells. This replication can be blocked by aphidicolin (Fig. 2, lane *h*), a specific inhibitor of DNA polymerase α , the polymerase that normally replicates both SV40 and the chromosomal DNA¹⁷.

It was an essential control for these experiments to determine whether the differences in replication levels we detected resulted from large difficulties in the transfection efficiencies of the cell lines treated. We examined this point extensively using several different assays. Perhaps the most relevant was a comparison by immunoprecipitation of the amount of large-T antigen synthesized when equal amounts of the non-replicating plasmid pMK16/ori⁻ were introduced into 293, HeLa and 143 cells (see Fig. 4*b*). The amount of large-T antigen synthesized in HeLa and 143 cells averaged 2.4- and 3.6-fold less, respectively, than in 293 as determined by densitometer scans of gels from two experiments and normalization for cell number. These small differences in transfection efficiency (and/or gene expression) cannot account for the large differences in replication observed. As an additional transfection control, we measured the number of stable hygromycin-resistant colonies resulting after transfection

of 293, JW-2, HeLa and 143 cells using a vector encoding resistance to hygromycin¹⁸. Our results indicated that these lines are comparable in transfection efficiency. This conclusion is supported by previous studies^{18–20}. Alwine²⁰ recently reported that 293 cells stabilize transfected DNA much better than other cell lines, although we have not been able to confirm this interpretation. Primary HEK cells (obtained from E. Major) take up DNA well, since they express large-T antigen efficiently following transfection with pMK16/ori⁻, as determined by immunoprecipitation (data not shown). Thus, we cannot explain the failure of HEK and lines such as 143 to replicate SV40 efficiently by an inability to take up DNA or express large-T antigen. Rather, a replication-permissive environment seems to exist in 293 cells.

We had originally considered that efficient SV40 replication in 293 might be caused by high levels of large-T antigen expression. The SV40 early promoter fused to the α -globin gene has been shown to be transcribed in 293 cells in an enhancer-independent manner¹¹. However, we were surprised to learn that D. Lewis and J. Manley (personal communication) had demonstrated that expression of large-T antigen from its own promoter on an SV40 vector is very low in 293 cells. On the other hand, these investigators also observed that SV40 early expression was easily detectable in 293 cells when a vector containing a defect in the origin of replication was used. In the light of our observation that SV40 replication is strong in 293, these results could suggest an antagonism between replication and gene expression. This interpretation was suggested by Botchan, whose laboratory had observed an apparent incompatibility between replication and transcription detected on SV40-based vectors encoding the herpes thymidine kinase gene (P. Robbins and M. Botchan, personal communication).

To test this hypothesis, we used indirect immunofluorescence to examine levels of large-T antigen in 293 cells transfected with replication-competent and replication-defective vectors. Figure 3*a* shows the immunofluorescence result for COS-7 cells. All the cells showed positive nuclear fluorescence; in contrast, no positive cells were detected when 293 was transfected with pJYMib or pMK16/SV40, another plasmid containing the whole SV40 genome¹⁶. Figure 3*b* shows a representative field of 293 cells transfected with pMK16/SV40. All the nuclei were dark; no bright nuclei were found in six independent plates examined. (The cytoplasmic background fluorescence is typical of 293 and occurs to the same degree in untransfected cells.) This result indicates that levels of large-T antigen must be very low or absent, even though vigorous pMK16/SV40 replication is readily detected in the same conditions.

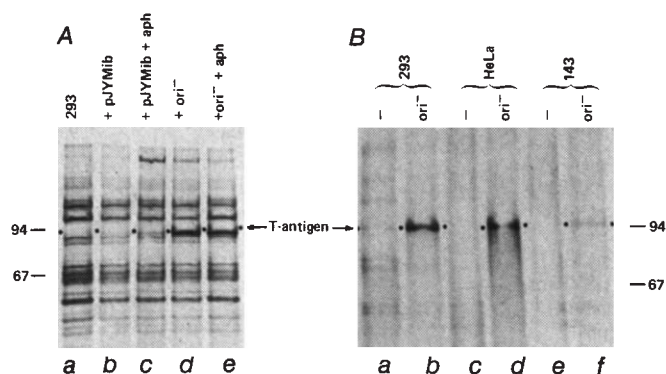


Fig. 4 Immunoprecipitation of large-T antigen from transfected cells. Approximately $1-5 \times 10^6$ cells were seeded onto 100-mm tissue culture dishes and transfected 24 h later with 10 μ g of the designated plasmid using calcium phosphate co-precipitation. Twenty hours after transfection, the cells were labelled with 100 μ Ci 35 S-methionine in methionine-deficient Dulbecco's modified Eagle's medium containing 0.5% fetal calf serum. After 24 h labelling, the cells were washed with PBS, lysed with 150 mM NaCl, 50 mM Tris pH 7.6 and 0.5% Nonidet P-40, and pelleted to remove cellular debris. The resulting supernatants were preabsorbed with IgG-sorb (Enzyme Center), then absorbed with an IgG-sorb-L19 monoclonal²⁷ anti-large-T antigen conjugate. The precipitates were washed with 150 mM NaCl, 50 mM Tris pH 7.6, 0.5% NP40 (A) or 500 mM NaCl, 50 mM Tris pH 7.6, 0.5% NP40 (B) to remove proteins which nonspecifically bind to the conjugates. The precipitates were boiled in 2% SDS, 20 mM Tris pH 6.8, 20% glycerol, 10% β -mercaptoethanol and run on 11% Laemmli SDS polyacrylamide gels²⁸. Large-T antigen is indicated by arrows in the figure and its position in the lanes is marked by a dot. Since only a small percentage of the cells were productively transfected, it was difficult to both retain the large-T antigen signal and completely remove labelled cellular background bands. The higher salt wash used in B was more successful in making large-T antigen the most strongly labelled band. The positions of 67,000 and 94,000 relative molecular mass markers (bovine serum albumin and phosphorylase b) are indicated. A, 293 cells were transfected with: a, a calcium phosphate precipitate with no DNA; b, c, pJYMib; d, e, pMK16/ori⁻. In c and e, 20 μ M aphidicolin was added to the culture medium 20 h after transfection. B, 2×10^6 293 cells (a, b), 4×10^6 HeLa cells (c, d) and 3×10^6 143 cells (e, f) were transfected with no DNA (a, c, e) or pMK16/ori⁻ (b, d, f). Relative amounts of large-T antigen were determined by densitometer scan of the gels, with normalization for cell number, and are reported in the text.

To see whether increased levels of large-T antigen could be obtained if replication were blocked, we performed immunofluorescence in 293 cells transfected with the replication-defective vector pMK16/ori⁻ (ref. 16), the large-T antigen-defective pMK16/SV80, and pMK16/Gm638, a large-T antigen mutant affecting amino-acid 457 and defective in replication (W. Gish and M. Botchan, personal communication). Immunofluorescence of cells transfected with these vectors produced bright positives, 5–10% of the cells displaying bright nuclear fluorescence with each of the mutants on each of three independent plates. Figure 3c shows an example of the positive nuclei observed with pMK16/ori⁻. This result is in contrast with that obtained using the replication-competent pMK16/SV40, where no positives were found (Fig. 3b). Furthermore, large-T antigen expression could be obtained from a replication-competent vector by blocking replication with aphidicolin, although these positives (Fig. 3d) were generally less intense than those obtained using mutant plasmids.

Our immunofluorescence results were corroborated by immunoprecipitation of large-T antigen 24 h after transfection (Fig. 4a). Addition of aphidicolin to cells transfected with pJYMib led to a threefold increase in the amount of large-T antigen synthesized. This result is remarkable considering that aphidicolin effectively blocks replication (Fig. 2h) so that the vector copy number is much lower than in the cells not treated with aphidicolin. The level of large-T antigen expression per

vector copy is therefore very much higher for the non-replicating situation. Large-T antigen levels for the pMK16/ori⁻ vector were two- or threefold higher than those obtained with the drug, paralleling the results obtained by us using immunofluorescence and by Lewis and Manley (see accompanying paper²¹) using cytosine arabinoside. The addition of aphidicolin to cells containing pMK16/ori⁻ produced no further rise in large-T antigen levels, showing that the increase is not due to presence of the drug *per se*, but rather is likely to be a result of the ability of aphidicolin to block replication.

We thus observe an inverse correlation between ability to replicate and levels of large-T antigen expression. This finding is in agreement with results obtained by Lewis and Manley in which large-T antigen expression is detectable in 293 only with replication-defective vectors²¹. Their RNA analysis indicates that the effect is operating at the transcriptional level²¹. Large-T antigen levels are known to be controlled in part by autoregulation^{22,23}. The antigen, by binding to the origin-early promoter region, inhibits its own transcription^{24,25}. However, we do not believe that autoregulation can completely explain our results because large-T antigen binding to the origin-promoter region is apparently normal for pMK16/ori⁻ (ref. 26) and in the presence of aphidicolin, yet in each of these situations, high levels of large-T antigen are observed. Instead, it appears that replication *per se* is antagonizing large-T antigen expression.

Our experiments show a requirement for large-T antigen in SV40 replication in 293 cells, yet immunofluorescence indicates that levels of the antigen expressed in 293 are very low. We suggest that vigorous SV40 replication can occur in 293 cells at marginal levels of large-T antigen because the cells are already primed for DNA synthesis. Large-T antigen has many functions. In its SV40-transforming function it induces a set of host genes involved in proliferation³. It is possible that these functions are subsumed by the E1 proteins, the adenovirus transforming functions.

We thank M. Botchan, P. Robbins and W. Gish for helpful discussions, vector DNA and 143 cells, M. Kreigler for instruction in immunofluorescence, J. Manley and D. Lewis for communication of unpublished results, P. Gallimore and J. Williams for JW-2 cells, E. Major for HEK cells, R. Tjian for a generous supply of L19 monoclonal antibody to large-T antigen, R. Schultz for XP cells, K. S. Greisen for assistance and J. Manley and R. Tjian for comments on the manuscript. J.S.L. was supported by the Leukemia Society of America, S.C. by a NIH training grant. This work was funded by grants to M.P.C. from the NCI and the Chicago Community Trust/Searle Scholars Program.

Received 11 June; accepted 25 July 1985.

- Calos, M. P., Lebkowski, J. S. & Botchan, M. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3015–3019 (1983).
- Lebkowski, J. S., DuBridge, R. B., Antell, E. A., Greisen, K. S. & Calos, M. P. *Molec. cell Biol.* **4**, 1951–1960 (1984).
- Toozie, J. *DNA Tumor Viruses* 2nd edn (Cold Spring Harbor Laboratory, New York, 1981).
- Lusky, M. & Botchan, M. *Nature* **293**, 79–81 (1981).
- Wigler, M. et al. *Cell* **16**, 777–785 (1979).
- Hirt, B. *J. molec. Biol.* **26**, 365–369 (1967).
- Bachetti, S. & Graham, F. L. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1590–1594 (1977).
- Major, E. O. & Matsumura, P. *Molec. cell Biol.* **4**, 379–382 (1984).
- Imperiale, M. J., Feldman, L. T. & Nevins, J. R. *Cell* **35**, 127–136 (1983).
- Green, M. R., Treisman, R. & Maniatis, T. *Cell* **35**, 137–148 (1983).
- Treisman, R., Green, M. R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7428–7432 (1983).
- Gaynor, R. B., Hillman, D. & Berk, A. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1193–1197 (1984).
- Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. J. *gen. Virol.* **36**, 59–72 (1977).
- Gluzman, Y. *Cell* **23**, 175–182 (1981).
- Whittaker, J. L., Byrd, P. J., Grand, R. J. A. & Gallimore, P. H. *Molec. cell Biol.* **4**, 110–116 (1984).
- Gluzman, Y., Frisque, J. & Sambrook, J. *Cold Spring Harb. Symp. quant. Biol.* **44**, 293–300 (1979).
- Huberman, J. A. *Cell* **23**, 647–648 (1981).
- Yates, J. L., Warren, N. & Sugden, B. *Nature* **313**, 812–815 (1985).
- Weeks, D. L. & Jones, N. C. *Molec. cell Biol.* **3**, 1222–1234 (1983).
- Alwine, J. C. *Molec. cell Biol.* **5**, 1034–1042 (1985).
- Lewis, E. D. & Manley, J. L. *Nature* **317**, 172–175 (1985).
- Tegtmeyer, P., Schwartz, M., Collins, J. K. & Rundell, K. J. *Virol.* **16**, 168–178 (1975).
- Reed, S. I., Stark, G. R. & Alwine, J. C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3083–3087 (1976).
- Tjian, R. *Cell* **13**, 165–179 (1978).
- Rio, D., Robbins, A., Myers, R. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5706–5710 (1980).
- Lepton, B. A., DeLucia, A. L. & Tegtmeyer, P. J. *Virol.* **49**, 9–13 (1984).
- Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. J. *Virol.* **39**, 861–869 (1981).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

Repression of simian virus 40 early transcription by viral DNA replication in human 293 cells

E. Diann Lewis & James L. Manley

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

The small DNA tumour virus simian virus 40 (SV40) has served as an excellent model for many studies on the mechanism and control of gene expression in eukaryotic cells. The SV40 early region produces two protein products. One product (large-T antigen) is known both to repress early viral transcription^{1,2} and to stimulate viral replication^{3,4} by binding to specific sites in the origin-promoter region⁵⁻⁷. The early promoter has several similarities to other RNA polymerase II promoters, for example, it possesses a TATA box, an upstream element⁸⁻¹⁰ and an enhancer^{11,12}. However, the SV40 early promoter differs from other known RNA polymerase II promoters in that the origin of viral DNA replication is embedded within it^{13,14}. Here we show that the SV40 early region is expressed at an extremely low level following its introduction into human 293 cells, contrasting with results observed in a large number of other cell lines¹⁵. We show further that the lack of expression is due to repression of transcription from the SV40 early promoter by viral DNA replication which occurs efficiently in 293 cells.

293 cells are human embryonic kidney cells that were transformed by and constitutively express the *E1a* and *E1b* genes of adenovirus-5 DNA^{16,17}. An *E1a*-encoded protein (or proteins) has been shown not only to be required for the positive regulation of transcription of other adenovirus genes^{18,19}, but also for activating²⁰, or, in some cases, repressing²¹, transcription of cellular genes. To examine expression of the SV40 early region in 293 cells, a recombinant plasmid containing the entire SV40 genome, pSVRI (Fig. 1a), was introduced into both HeLa and 293 cells by calcium phosphate co-precipitation²² and its expression was monitored by indirect immunofluorescent²³ visualization of large-T antigen in cells fixed 48 h following transfection (Table 1). As expected, efficient expression of large-T antigen was observed in HeLa cells; however, no indication of its expression could be detected in 293 cells. In contrast, the plasmid pφ4-SVAe, which contains the adenovirus late promoter fused to SV40 large-T antigen-encoding sequences (see Fig. 1a), gave rise to efficient expression of large-T antigen in 293 cells as well as in HeLa cells²⁰. The observed lack of large-T antigen expression in pSVRI-transfected 293 cells was not unique to pSVRI. We have tested several different plasmids containing intact SV40 early regions, including pSTER (Fig. 1), and all failed to express large-T antigen in 293 cells, as judged by indirect immunofluorescence.

To determine whether mutations in the promoter-origin region might 'activate' the early promoter in 293 cells, we first tested a mutant, pSVRIori⁻, which contains a small insertion at the origin of replication, defining a *Bgl*I restriction site (see Fig. 1). Because this mutation has virtually no effect on early gene expression in several other cell types^{24,25}, although preventing viral DNA replication in permissive monkey cells²⁵, we were surprised to find that 293 cells transfected with pSVRIori⁻ produced significant quantities of large-T antigen (Table 1). In fact, the percentage of cells expressing the antigen was roughly the same as in cells transfected with pφ4-SVAe. In contrast, HeLa cells behaved very like other cells in their response to the origin mutation; that is, no effect was detectable (Table 1). To determine whether the observed activation was unique to pSVRIori⁻, we tested several other mutants (obtained from Y. Gluzman) possessing deletions at the *Bal*I restriction site²⁶ for large-T antigen production in both 293 and HeLa cells. The results (Table 1) are similar to what was observed with pSVRIori⁻; large-T antigen was expressed efficiently in both cell types.

Table 1 Expression of large-T antigen from SV40 plasmids in HeLa and 293 cells by indirect immunofluorescence

Plasmid	% T-positive cells	
	HeLa	293
pφ4-SVAe	31	10.5
pSVRI	38	<0.005§
pC6K1*	20.3	15.2
pC2	29	15.5
pC11	32.2	10.5
pSVRIori ⁻	40	11.8
p8-4†	40.7	14.4
p6-1	15	7.1
p3-20	35.8	3.6
pSVRI + AraC‡		0.4
pSVRI + HU		1.5
pSVRIori ⁻ + AraC		4.3
pSVRIori ⁻ + HU		4.4

Supercoiled plasmid DNA (6 µg) was transfected onto 4×10^5 cells (either 293 or HeLa) on a 60-mm plate by calcium phosphate co-precipitation²². Four hours after the addition of the DNA precipitate to 293 cells, the DNA and medium were removed and the cells were glycerol shocked with 10% glycerol for 1.5 min²⁹. The cells were washed with TBS⁴⁰, fresh Dulbecco's modified Eagle's medium (Gibco) with 10% fetal bovine serum (Gibco) was added and the cells were incubated for a further 44 h. The HeLa cells were glycerol shocked after 18 h of contact with the DNA precipitate, and incubated a further 30 h. For the indirect immunofluorescence experiments, cells that had been grown on coverslips were fixed in 3.7% formaldehyde for 15 min and treated with Nonidet P-40 for 1 min. Large-T antigen was detected by first adding a mouse monoclonal antibody to large-T antigen (Pab 416)⁴¹ and then staining with fluorescein-conjugated goat anti-mouse second antibody (Cappel). Cells were visualized with a Zeiss fluorescent microscope. These results represent the average of several separate experiments.

* pC6K1 has three base changes affecting three amino acids¹⁸. pC2 has a mutation of one amino acid and pC11 has base changes affecting two amino acids of large-T antigen⁴⁹.

† p8-4, p6-1 and p3-20 have 4 base pairs (bp), 6 bp and 58 bp deletions, respectively²⁶.

‡ Four hours before the addition of the DNA precipitate, the medium on the cells was changed to fresh medium with either 2 µg ml⁻¹ of cytosine arabinoside (AraC) or 1 mM hydroxyurea (HU). The transfection protocol described above was followed except that medium supplemented with the appropriate drug was added following the glycerol shock. Although the percentage of cells expressing large-T antigen was low, significant numbers of positive cells were detected in each of five separate experiments.

§ 0.005% means that no T-positive cells were ever detected. This experiment has been repeated more than 20 times.

To characterize further the block to large-T antigen synthesis in 293 cells, we analysed total cytoplasmic RNA obtained from transfected 293 and HeLa cells by quantitative 5' S₁ nuclease mapping. The results (Fig. 2) are consistent with the immunofluorescence analysis. Although readily detectable amounts of SV40 early RNA, with the correct 5' ends, were detected in the RNA obtained from pSVRI-transfected HeLa cells, virtually no specific RNA was found in samples obtained from 293 cells transfected with that DNA. However, both HeLa and 293 cells that had been transfected with pSVRIori⁻ produced significant quantities of SV40-specific RNA, with a single 5' end (shifted several nucleotides upstream by the insertion mutation).

The position of these mutations suggested at least two possible explanations for our results. First, as *Bgl*I mutations are within large-T antigen binding site 2 (see Fig. 1b), it was possible that they might affect large-T antigen-mediated autoregulation. By this model, very low levels of large-T antigen would be synthesized in pSVRI-transfected 293 cells which would, however, be sufficient to block further early transcription in a form of 'super' autoregulation, so that very low levels of large-T antigen protein and messenger RNA would accumulate in the transfected cells. We considered this explanation unlikely for several reasons, including the fact that the major large-T antigen binding site

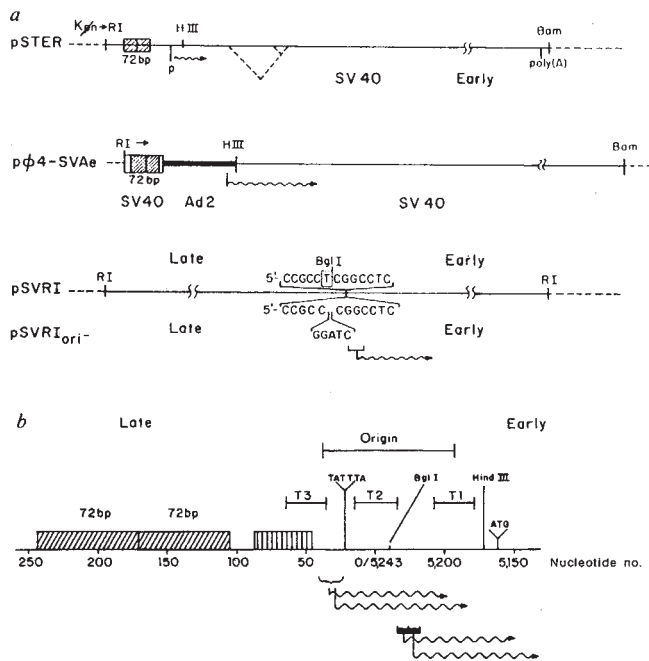


Fig. 1 *a*, Plasmid constructs. pSTER was constructed from pSVAdL (D. S. Grass and J.L.M., manuscript in preparation), which contains an *Eco*RI linker at the unique *Kpn*I site (at nucleotide 294) of SV40. The *Eco*RI (at nucleotide 294) to *Bam*HI (at nucleotide 2,533) fragment of SV40 containing the entire early region was inserted into the corresponding sites in pBR322. The plasmid pφ4-SVAe contains the 72-bp repeats of SV40¹¹, 438 bp of the adenovirus-2 (Ad2) late promoter from -405 to +33 (relative to the Ad2 transcription start site at +1) attached to the protein coding sequence of the SV40 early region at the *Hind*III site at nucleotide 5,171 of SV40 (for details of construction, see ref. 42). The construction of pSVRI and pSVRI_{ori}⁻ are described in ref. 25. These plasmids contain the entire SV40 genome cloned into pBR322 at the respective *Eco*RI sites. All restriction enzymes used were from New England Biolabs. pBR322 sequences are represented by dashed lines. SV40 sequences are represented by a narrow line and hatched boxes, while the Ad2 late promoter is represented by a thick line. *b*, Enlargement of the SV40 origin of replication-promoter region. Cross-hatched boxes, 72-bp enhancers; hatched box, G+C-rich segment; solid bracket, heterogeneous early transcription starts; narrow bracket, early region transcription starts seen at late times in an SV40 infection. T1, T2, T3 are large-T antigen binding sites. The numbering system of Buchman *et al.*⁴³ is used. This figure was derived from one in ref. 44.

(site 1) implicated in autoregulation²⁷ is intact in pSVRI_{ori}⁻. A second possibility was suggested by the observation that all the mutations that activated large-T antigen expression are defective in DNA replication. This raised the possibility that only plasmids that could not replicate their DNA would be able to express large-T antigen efficiently in 293 cells, that is, replication and early transcription are incompatible.

To distinguish between these two alternatives, we analysed a series of recombinant plasmids (obtained from Y. Gluzman) containing point mutations within the large-T antigen-encoding sequences²⁸. Large-T antigen from one mutant, pC6, is both defective in catalysing viral DNA replication and also fails to bind specifically to viral DNA⁷. Since autoregulation is effected by specific DNA binding, this mutant is presumably defective in autoregulation. Two other mutants, pC2 and pC11, are also defective in DNA replication, but retain the ability to bind specifically and efficiently *in vitro* to the viral origin (to both sites 1 and 2)⁷, and are presumably unaffected in their ability to autoregulate.

pC2, pC6 and pC11 DNAs were used to transfect both HeLa and 293 cells. The results obtained from indirect immunofluorescence assays were clear. Although large-T antigen expression in HeLa cells was essentially unaffected by any of these mutations, all three plasmids expressed large-T antigen efficiently in

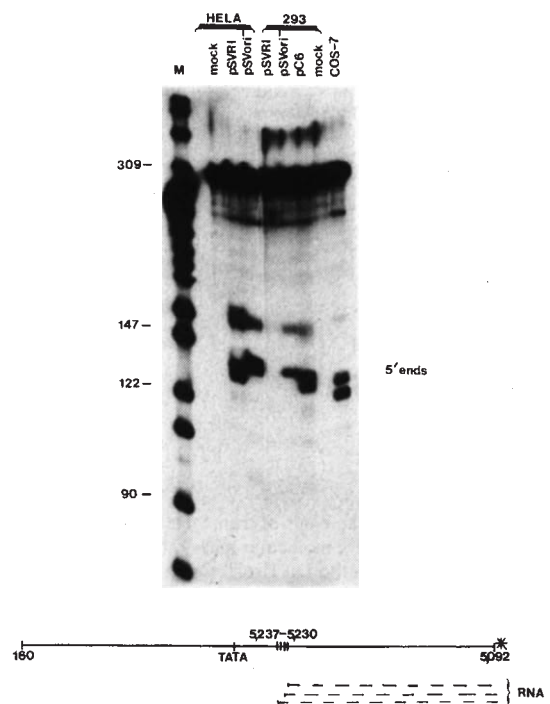


Fig. 2 S₁ analysis of SV40-specific RNA from 293 or HeLa cells after transfection with the plasmid DNA. Lane M, relative molecular mass markers ($\times 10^{-3}$). The sample analysed in the 'COS-7' lane contained total cytoplasmic RNA from an SV40-transformed monkey cell line⁴⁵; the SV40 DNA used to make the cell line contains a 6-bp deletion at the *Bgl*II site, which causes the transcription start sites to shift downstream by 6 bp relative to wild type. The asterisk in the diagram below signifies the labelled end of the probe.

Methods. 25 μ g of each plasmid DNA was used to transfect 1.8×10^6 cells on a 150-mm plate. 48 hours after the addition of DNA to cells, the medium was removed, 5 ml of cold TSM (150 mM NaCl, 10 mM Tris-HCl pH 7.5, 3 mM MgCl₂) was added, and the cells were scraped off the plate. The plate was washed with 2 ml more TSM and rescraped. The cells were pelleted and washed three times with 7 ml of cold TSM for each wash at 4 °C. After the final wash, the cells were resuspended in 0.5 ml TSM which was brought to 0.6% Triton X-100 (Sigma) and incubated for 5 min on ice. The nuclei were pelleted and the supernatant was taken and designated total cytoplasmic RNA. This RNA was extracted three times with phenol and chloroform, precipitated with 2½ volumes of ethanol and stored in 0.2% sarkosyl, 1 mM MEDTA at -20 °C. 30 μ g of total cytoplasmic RNA was hybridized for 4 h at 43 °C to 10 ng of denatured, double-stranded DNA probe labelled at the 5' end. The DNA probe used was the 311-bp *Bst*NI fragment of SV40 extending from nucleotide 160 to 5,902. Nuclease S₁ digestion⁴⁶ was at 43 °C for 30 min with 8×10^3 U ml⁻¹ of enzyme (Boehringer Mannheim). The digestion products were electrophoresed through an 8% polyacrylamide-8 M urea sequencing-type gel⁴⁷. The DNA size marker was pBR322 digested with *Hpa*II and 5' end-labelled with T4 polynucleotide kinase (New England Biolabs).

293 cells, in striking contrast to 293 cells transfected with plasmids encoding a wild-type large-T antigen (Table 1). These results were confirmed for one of the mutants (pC6) at the RNA level by quantitative S₁ nuclease mapping (Fig. 1) and show that significant quantities of SV40-specific early RNA, with the same 5' ends as wild-type RNA (in HeLa cells), were produced in 293 cells, and suggest that super autoregulation is not the basis for the absence of large-T antigen expression observed in 293 cells transfected with plasmids containing a wild-type SV40 early region. The results are, however, consistent with a strong repression of early transcription resulting from large-T antigen-dependent, origin-specific DNA replication in 293 cells.

It was therefore of interest to learn that Calos and co-workers had recently obtained evidence that SV40 DNA does indeed replicate efficiently in 293 cells (see accompanying paper)²⁹. To

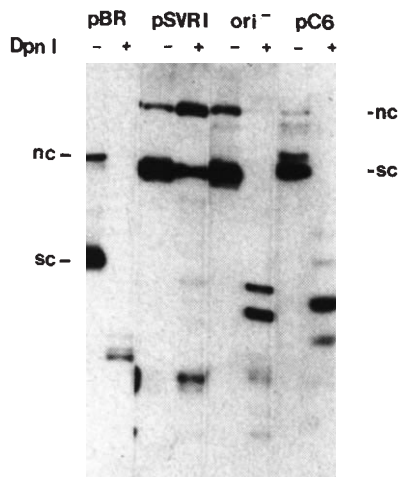


Fig. 3 Southern blot analysis of transfected DNA. 48 hours after transfection, low relative molecular mass DNA was extracted from 60-mm plates by the method of Hirt³⁰. One-twentieth of each DNA sample was used for digestion with 8 units of *DpnI* (Bethesda Research Labs) at 37°C for 5 h. Equal amounts of each sample were electrophoresed through a 1% agarose (Sigma) gel, transferred to nitrocellulose (Schleicher & Schell)⁴⁸ and hybridized to nick-translated pSVRI probe. sc, supercoiled, nc, nicked circles, ori⁻, pSVRIori⁻.

test whether this was so in our system, Hirt supernatants³⁰ were prepared from cells transfected with different plasmids and aliquots were incubated with the restriction endonuclease *DpnI* (this enzyme only cleaves DNA grown in bacteria and therefore methylated at the adenosine residue in its recognition signal). DNA samples were then analysed by the method of Southern³¹, and plasmid DNA was detected by hybridization with nick-translated³² pSVRI DNA. Only DNA that replicated in the animal cell will remain intact (that is supercoiled or nicked circles) after treatment with *DpnI*; non-replicated DNA will be digested to small fragments. Significant quantities of *DpnI*-resistant (replicated) DNA were present in samples from cells that had been transfected with pSVRI (Fig. 3). Plasmid DNA from cells transfected with pBR322, pSVRI ori⁻ or pC6 were digested to completion. We conclude that pSVRI does indeed replicate in 293 cells, and that this replication requires an intact origin of replication and a functional large-T antigen.

Because mutations that block replication activate the synthesis of large-T antigen, we wondered whether chemical inhibitors of DNA replication might allow early gene expression to occur in 293 cells. To test this hypothesis, we carried out transient expression assays in the presence or absence of cytosine arabinoside or hydroxyurea, known inhibitors of SV40 (and cellular) DNA replication, and monitored large-T antigen expression by indirect immunofluorescence. The results (Table 1) show that large-T antigen-positive cells could indeed be detected following transfection of 293 cells in the presence of either drug. Although the percentage of cells expressing large-T antigen was low (caused at least in part by a toxic effect of the drugs; large-T antigen expression with pSVRI ori⁻ was in fact reduced by the drugs), significant numbers of positive cells were detected in each of four separate experiments. This contrasts sharply with results obtained in the absence of the drugs, where we have never detected a large-T antigen-positive cell. That the observed activation of large-T antigen expression was a result of the inhibitory effect of the drugs on DNA replication is supported by the observations of Lebkowski *et al.*²⁹, who obtained results essentially identical to ours, except that in their experiments DNA synthesis was blocked with aphidicolin. These results strongly suggest that viral DNA replication represses early transcription in 293 cells.

This situation seems paradoxical: early transcription is necessary to produce large-T antigen, which is in turn required for viral DNA replication to occur. Thus, we propose that small

amounts of large-T antigen are synthesized in transfected 293 cells, and that this amount is sufficient to catalyse DNA replication, which then further represses early transcription. Our results thus raise the possibility that either much lower levels of large-T antigen than previously thought are sufficient to bring about DNA replication and/or that its role in DNA replication in 293 cells is different from its action in permissive cells.

An interplay between DNA replication and transcription is not without precedent. For example, in several prokaryotic and eukaryotic viruses, shifts from 'early' to 'late' gene expression are associated with the onset of viral DNA replication. Although the so-called late promoter of adenovirus is active at very early times after infection^{33,34}, the entire late transcription unit is not expressed until late times. DNA replication seems to be required for the RNA polymerase to traverse the entire transcription unit³⁵; that is, changes in the DNA template brought about by replication seem to affect its interaction with RNA polymerase.

Several other examples suggest an inverse correlation between DNA replication and transcription. For example, transcription from the human α -globin promoter is increased ~50-fold in transient expression assays when the gene is contained on replicating, relative to non-replicating, plasmids; however, transcription from the β -globin promoter is not affected by replication³⁶. Additionally, DNA replication is necessary to bring about repression of silent mating type loci in yeast³⁷.

Why are 293 cells proficient at SV40 DNA replication? The notion that an adenovirus *E1a* and/or *E1b* gene product is involved is strengthened by the findings²⁹ not only that human embryonic kidney (HEK) cells (the cell type from which 293 cells are derived) are unable to replicate SV40 DNA, but also that another independent *E1a* plus *E1b*-transformed HEK cell line (JW-2) also supports efficient SV40 DNA replication. We suggest that 293 cells have higher concentrations of factors required to replicate SV40 DNA. For example, the concentration of topoisomerase I is 10-fold greater in 293 cells than in HeLa cells³⁸. This may result from the known ability of *E1a*-encoded proteins to activate gene expression.

We thank C. Prives and M. Botchan for stimulating conversations during this work, C. Prives for her comments on the manuscript, M. Calos for communicating her results to us before publication, G. Blanck for supplying nick-translated pSVRI in the early stages of these experiments, and Y. Gluzman for providing plasmids. Technical assistance was by H. Thomas and R. Mathews. This work was supported by NIH grants GM28983 and CA33620.

Received 8 May; accepted 24 July 1985.

1. Tegtmeyer, P., Schwartz, M., Collins, J. K. & Rundell, K. *J. Virol.* **16**, 168-178 (1975).
2. Reed, S. I., Stark, G. R. & Alwine, J. C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3083-3087 (1976).
3. Tegtmeyer, P. *J. Virol.* **10**, 591 (1972).
4. Chou, J. Y., Avila, J. & Martin, R. G. *J. Virol.* **14**, 116-124 (1974).
5. Hansen, U., Tenen, D. G., Livingstone, D. M. & Sharp, P. A. *Cell* **27**, 603-612 (1981).
6. Rio, D., Robbins, A., Myers, R. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5706-5710 (1980).
7. Prives, C., Covey, L., Scheller, A. & Gluzman, Y. *Molec. cell. Biol.* **3**, 1958-1966 (1983).
8. Manley, J. L. *Prog. Nucleic Acids Res. molec. Biol.* **30**, 195-244 (1983).
9. Benoist, C. & Chambon, P. *Nature* **290**, 304-309 (1981).
10. Everett, R. D., Baty, D. & Chambon, P. *Nucleic Acids Res.* **11**, 2447-2464 (1983).
11. Gruss, P., Dhar, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 943-947 (1981).
12. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047-6068 (1981).
13. Danna, K. S. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3097-4001 (1972).
14. Bergsma, D. J., Olive, D. M., Hartzell, S. W. & Subramanian, K. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 381-385 (1982).
15. Khoury, G., Byrne, J. C., Takemoto, K. K. & Martin, M. A. *J. Virol.* **11**, 54-60 (1973).
16. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. J. *gen. Virol.* **36**, 59-72 (1977).
17. Aiello, L., Guilfoyle, R., Huebner, K. & Weinmann, R. *Virology* **94**, 460-469 (1979).
18. Jones, N. & Shenk, T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3665-3669 (1979).
19. Berk, A. J., Lee, F., Harrison, T., Williams, J. & Sharp, P. A. *Cell* **17**, 937-944 (1979).
20. Green, M. R., Treisman, R. & Maniatis, T. *Cell* **35**, 137-148 (1983).
21. Schrier, P. I., Bernards, R., Vaessen, R. J., Houweling, A. & van der Eb, A. J. *Nature* **305**, 771-775 (1983).
22. Wigler, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 1373-1376 (1979).
23. Verderame, M., Alcorn, D., Egnor, M., Smith, K. & Pollack, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6624-6628 (1980).
24. Lewis, E. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7065-7069 (1983).
25. Chen, S. *et al. J. Virol.* **48**, 492-502 (1983).
26. Gluzman, Y., Sambrook, J. F. & Frisque, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3898-3902 (1980).
27. DiMaio, D. & Nathans, D. *J. molec. Biol.* **156**, 531-548 (1982).
28. Gluzman, Y., Davison, J., Oren, M. & Winocour, E. *J. Virol.* **22**, 256-266 (1977).
29. Lebkowski, J. S., Clancy, S. & Calos, M. P. *Nature* **317**, 169-171 (1985).
30. Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).

31. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
32. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-252 (1977).
33. Shaw, A. R. & Ziff, E. B. *Cell* **22**, 905-916 (1980).
34. Akusjarvi, G. & Persson, H. *Nature* **292**, 420-426 (1981).
35. Thomas, G. P. & Mathews, M. B. *Cell* **22**, 523-533 (1980).
36. Treisman, R., Green, M. R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7428-7432 (1983).
37. Miller, A. M. & Nasmyth, K. A. *Nature* **312**, 247-251 (1984).
38. Chow, K.-C. & Pearson, G. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2247-2251 (1985).
39. Frost, E. & Williams, J. *J. Virol.* **91**, 39-50 (1978).
40. Kimura, G. & Dulbecco, R. *Virology* **49**, 394-403 (1972).
41. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861-869 (1981).
42. Lewis, E. D., Fu, X.-Y. & Manley, J. L. *ICN-UCLA Symp. molec. Biol. Dev.* **19**, 351-360 (1984).
43. Buchman, G., Burnett, L. & Berg, P. in *DNA Tumor Viruses* 2nd edn (ed. Tooze, J.) 799-841 (Cold Spring Harbor Laboratory, New York, 1981).
44. Tooze, J. (ed.) *DNA Tumor Viruses* 2nd edn (Cold Spring Harbor Laboratory, New York, 1981).
45. Gluzman, Y. *Cell* **23**, 175-182 (1981).
46. Berk, A. J. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1274-1278 (1978).
47. Maxam, A. M. & Gilbert, W. *Methods Enzym.* **65**, 499-561 (1980).
48. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
49. Manos, M. M. & Gluzman, Y. *Molec. cell. Biol.* **4**, 1125-1133 (1984).
50. Lewis, E. D. and Manley, J. L. *Molec. cell. Biol.* (in the press).

Chromosomal localization of a unique gene by non-autoradiographic *in situ* hybridization

J. E. Landegent*, N. Jansen in de Wal*, G.-J. B. van Ommen†, F. Baas‡, J. J. M. de Vijlder‡, P. van Duijn* & M. van der Ploeg*

* Department of Histochemistry and Cytochemistry, and

† Department of Anthropogenetics, Sylvius Laboratories, State University, 2333 AL Leiden, The Netherlands

‡ Department of Pediatric Endocrinology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands

During the past few years, several methods have been developed for the detection of specific nucleic acid sequences by *in situ* hybridization using non-radioactive labels such as fluorochromes, cytochemically detectable enzymes and electron-dense markers¹⁻⁴. These methods are preferable to autoradiography in terms of speed of performance and topological resolution. Their limited sensitivity, however, has so far restricted their use to the detection of repeated sequences. Here we report single gene detection with a procedure using 2-acetylaminofluorene (AAF)-modified probes, immunoperoxidase cytochemistry and reflection-contrast microscopy. We confirmed the autoradiographic data on the localization of the human thyroglobulin (*Tg*) gene to the distal end of the long arm of chromosome 8 (ref. 5). A mixture of cosmid cHT2-derived subclones of the 3' part of the *Tg* gene, 22.3 kilobase pairs (kbp) in total, was used as a hybridization probe. This procedure can be used to map other unique sequences, if genomic clones are available from which clones with an appropriate amount of inserts can be isolated.

In non-autoradiographic procedures, labels are either coupled directly to the probe^{6,7} or introduced via specific antibodies after hybridization with haptenized probes^{2,8}. The procedure described here is an example of the latter, in which the nucleic acids are chemically modified with *N*-acetoxy-2-acetylaminofluorene (*N*-AcO-AAF). This compound reacts mainly at the C-8 position of guanine, yielding *N*-(guanin-8-yl)-AAF adducts⁹. Using anti-AAF antibodies and a fluorochrome-conjugated second-layer antiserum, we could localize high- and middle-repetitive sequences on mammalian metaphase chromosomes¹⁰. Recent developments in the visualization of *in situ* hybridization results have prompted us to investigate whether this procedure could be used for the detection of unique sequences. The *Tg* gene was used, because the gene is well characterized^{11,12} and recently localized by more conventional methods such as Southern blot analysis of human-rodent somatic cell hybrids with human-specific chromosomal losses and *in situ* hybridization with titrated probes⁵.

To establish a non-autoradiographic assignment of the *Tg* gene, hybridizations *in situ* with AAF-labelled probes were

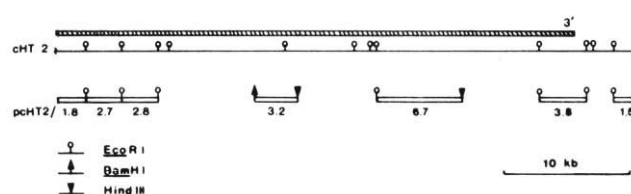


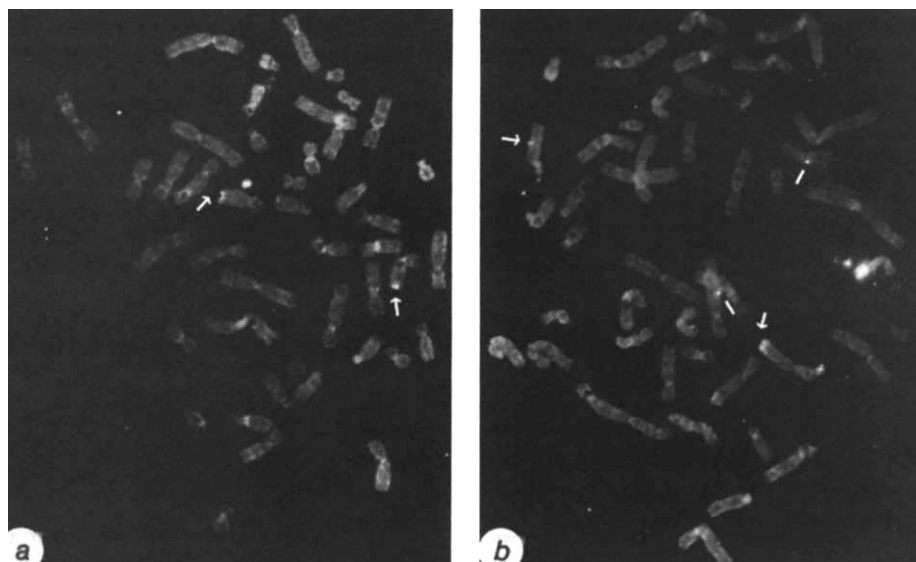
Fig. 1 Single-copy probes of the human thyroglobulin gene used for *in situ* hybridization (results from ref. 12). The upper bar shows the position of the 3' end of the *Tg* gene with respect to cosmid cHT2 (line). The complete restriction maps for *Eco*RI, *Bam*HI and *Hind*III used for subcloning in pAT 153 are also indicated. The single-copy subclones are depicted as pCHT2/ followed by the length of their insert (lower bar).

performed. A mixture of seven single-copy subclones derived from cosmid cHT2 was used to eliminate highly repeated sequences in genomic cosmid DNA. This cosmid contains 45 kbp, 38 kbp of which comprise the 3' end of the *Tg* gene¹¹. The subclones cover a total of 22.3 kbp, including 3 of the 5 exons in this region (0.55 kbp in total), ~20 kbp of intronic DNA and 2 kbp of 3'-flanking sequences¹² (Fig. 1). The DNA of these recombinant plasmids was sonicated to a single-stranded (ss) length of ~400 bases and modified with *N*-AcO-AAF to a substitution degree of about 5%¹⁰. The probe mixture was hybridized to metaphase spreads of cultured human lymphocytes and detected by means of an indirect immunoperoxidase reaction.

The diaminobenzidine (DAB) precipitates, deposited at the chromosomal sites of hybridization, were observed with reflection-contrast microscopy^{13,14}. This type of microscopy allows the visualization of reflection, rather than absorption of incident light by the dye precipitate. The light is reflected because of a critical difference in the refractive indices of the stained structure and the embedding medium, a phenomenon termed anomalous dispersion¹⁵. Because the DAB precipitate does not dissolve or recrystallize, these preparations are permanent and can be evaluated under the microscope several times without loss of detail. In addition to the specific hybridization signal, a low level of overall chromosome staining can be observed. This eliminates the need for a supplementary total DNA counterstain (Fig. 2). The applied combination of *in situ* hybridization and visualization also produces a banding pattern in the chromosomes resembling R-bands¹⁶. Although the banding is not as detailed as can be obtained with, for example, a routine G-banding procedure¹⁷, the chromosomes can be karyotyped. Direct microscope evaluation of the chromosomes is possible but a higher contrast is obtained on photographs taken from the metaphases. Standard banding protocols, when tested, had the disadvantage that specific hybridization signals could no longer be distinguished from the other staining results.

The exact localization of the *Tg* gene was determined by analysing the distribution of DAB reflection signals from 40 metaphase spreads of a representative slide. The scoring was performed from photographs of randomly selected metaphases by two observers who were unaware of the gene involved. Different types of signals can be seen (Fig. 2), as dust particles or other contaminants present in the incubation solutions or immersion oil can cause reflection of the incident light. These background signals can be readily distinguished from DAB reflection by a few criteria, such as a different colour (blue instead of yellow), a halo effect, or a difference in focus plane between the signal and the chromosomes. Of the total of 99 chromosomal sites labelled, 36 (36%) were located at the end of the long arm of chromosome 8 and the remainder were randomly distributed over the chromosomes (Fig. 3). More illustrative, 45% of the individual metaphases showed one or more chromatids of chromosome 8 containing a signal at the q24 band. In several instances, all four chromatids of a single metaphase spread were positive (Fig. 2a). The fact that not every possible specific hybridization site reacts, can at least partly be

Fig. 2 Non-autoradiographic *in situ* hybridization of single-copy human *Tg* gene probes. Two examples of metaphases, which show in addition to the specific signals a weak overall DNA staining and an R-banding pattern, are shown. *a*, Each of the four chromatids is labelled at the end of the long arm of chromosome 8 (arrows). *b*, Both chromatids of one chromosome 8 are positive. In the latter metaphase, a DAB signal, which was also scored (Fig. 3), is indicated as well. A few background signals showing halo effects are marked by bars. Original magnification $\times 1,000$.



Methods. A mixture of seven recombinant plasmids, derived from the 3'-end *Tg* cosmid clone cHT2 (see Fig. 1), was used as a probe. These plasmids, containing 22.3 kbp of unique fragments of the gene in total, were modified with AAF and hybridized to human metaphase preparations. The hybridized probes were

detected by an indirect immunoperoxidase/diaminobenzidine reaction in combination with reflection-contrast microscopy. Chromosome preparations of cultured human peripheral blood lymphocytes were made according to conventional methods. Amounts of 1 μ g DNA of each probe were pooled, sonicated to a ss length of ~ 400 bases, and labelled with *N*-acetoxy-2-acetylaminofluorene (*N*-AcO-AAF) to a substitution degree of about 5% (ref. 10). After purification the nucleic acids were dissolved at a concentration of 15 μ g μ l⁻¹ in the hybridization mixture, containing salmon sperm DNA (750 μ g μ l⁻¹), formamide (50%, v/v) and 2 $\times$ SSC (0.3 M NaCl, 30 mM Na-citrate, pH 7.0). Five μ l of this mixture were applied per microscopic slide and hybridized for 18 h at 37 °C. The slides were denatured for 5 min in 0.07 M NaOH in 70% ethanol and treated for 15 min at 37 °C with proteinase K (10 μ g per 100 ml 20 mM Tris-HCl, pH 7.5, 2 mM CaCl₂) prior to the hybridization. Immunocytochemical detection occurred through subsequent incubation with a rabbit anti-AAF serum (diluted 1:500 in phosphate-buffered saline (PBS) containing 2% normal serum and 0.05% Tween 20; SERVA), and a peroxidase-conjugated goat anti-rabbit serum (DAKO, diluted 1:50 in the same buffer) for 1 h at room temperature each. The slides were washed 3 times for 10 min in PBS between and after the incubations. The preparations were then stained with 0.01% 3,3'-diaminobenzidine tetrahydrochloride (Merck) and 0.01% H₂O₂ in 0.05 M Tris-HCl, pH 7.6 for 20 min at room temperature in the dark. After rinsing with PBS (3 times for 10 min), the preparations were dehydrated, air-dried and visualized by reflection-contrast microscopy¹⁴.

explained by loss of chromosomal DNA. We have found that under optimal conditions this loss still equals $\sim 30\%$ of the DNA initially present (Raap *et al.*, in preparation). In control experiments, which included both unrelated AAF-modified probes and unrelated primary antibodies on hybridized slides, no significant labelling was observed.

This 'statistical localization' is typical of autoradiographical chromosomal mapping of single-copy genes^{18,19}. The reported assignment of the *Tg* gene to the 8q24 band *in situ*⁵, obtained by the use of a similar set of probes, required the observation of a significant number of silver grains in 50 spreads. However, those results could only be evaluated after exposure times varying from 2 to 3 weeks, whereas our slides can be viewed the next day. Another disadvantage of autoradiography is the location of most of the silver grains near the hybridized probe, rather than on the chromosome itself, because of the thickness of the emulsion.

In its present form the technique allows for the first time the non-autoradiographic detection of single-copy genomic sequences. In terms of slide preparation before hybridization, two elements are significant: the addition of alcohol to the denaturation mixture, which is important for the preservation of DNA sequences¹⁰, and the application of proteinase K in enhancing accessibility for the antibodies²⁰. The use of a set of unique recombinant plasmids as a hybridization probe and reflection-contrast microscopy may further account for the sensitivity obtained. This method of visualizing DAB precipitates favourably compares with absorption microscopy¹⁴; DAB signals on these preparations (both the banding pattern and the hybridization signals), for example, are not visible with the latter method.

Other unique genes can be localized by the strategy of using flanking sequences in combination with coding fragments. The necessary subcloning of single-copy segments from large genomic clones is not a disadvantage and usually required for gene characterization. The precise lower detection limit of the

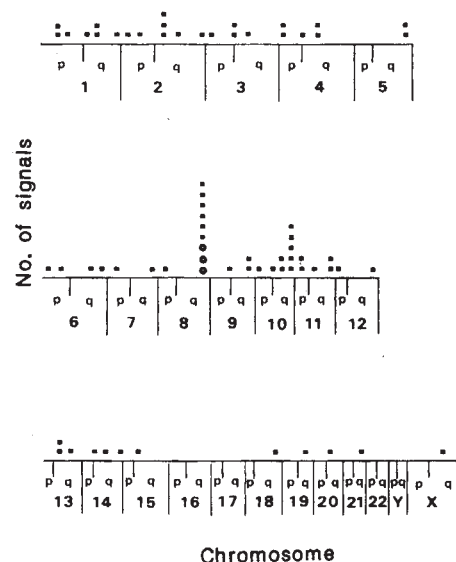


Fig. 3 Distribution of DAB reflection signals over metaphase chromosomes. Human chromosome preparations were hybridized with the AAF-labelled *Tg* probe mixture, visualized by means of the immunoperoxidase/reflection-contrast combination described in Fig. 2 legend and 40 randomly selected metaphases were analysed. The chromosomes were karyotyped by the R-banding pattern, which is provided by the procedure itself. The DAB reflection signals were scored and are marked as filled squares (■) above the schematic representation of the corresponding chromosome. Ten signals are indicated as filled circles (●). Thirty-six chromatids of chromosome 8 exhibited a signal at the distal part of the long arm (8q24 band), representing 36% of all labelled sites (99 in total) throughout the chromosome complement. This unambiguously confirms earlier autoradiographic data⁵ on the localization of the *Tg* gene.

procedure has not yet been determined, although it may be lower than the 22.3-kbp genomic DNA used here. Furthermore, the sensitivity of the method can still be improved at various levels of the procedure, for example, by optimization of the hybridization efficiency. (The application of pure ssRNA, obtained by *in vitro* asymmetrical transcription of recombinants with SP6 RNA polymerase²¹ has been reported to result in an eightfold increase of the *in situ* hybridization signals²².) Improvements also seem possible at the level of immunocytochemical detection. The use of the streptavidin biotinylated-peroxidase system in combination with AAF-modified probes has proved to be advantageous over an indirect peroxidase reaction²³. With such refinements, the localization of smaller single-copy fragments should be feasible. The speed of detection and the high resolution of the technique should greatly favour its applicability for the localization of structural genes or breakpoints in chromosomal rearrangements. In particular, the determination of the orientation of loci within a chromosome²⁴ would benefit from a non-autoradiographic approach.

We thank Dr P. Borst for his interest in our work, Dr R. A. Baan for N-AcO-AAF and the anti-AAF antisera and Drs G. C. Beverstock and F. M. Pet for karyotyping the chromosomes. This work was supported by FUNGO, Foundation of Medical Scientific Research in the Netherlands, contract 13-54-32.

Characterization of 64-, 123- and 182-base-pair exons in the mouse $\alpha 2(\text{IV})$ collagen gene

Markku Kurkinen*, Michael P. Bernard*†, Denise P. Barlow‡§ & Louise T. Chow§

Departments of Biochemistry*, Obstetrics and Gynecology†, University of Medicine and Dentistry of New Jersey, Rutgers Medical School, Piscataway, New Jersey 08854, USA
‡ Imperial Cancer Research Fund, Mill Hill, London NW7 1AD, UK
§ Department of Biochemistry, University of Rochester, School of Medicine and Dentistry, Rochester, New York 14642, USA

Genes encoding types I, II and III collagens (fibrillar collagens) contain many discrete-size exons, most of them 54 base pairs (bp) long, in addition to the 45-, 99-, 108- and 162-bp exons¹⁻⁶. It has been suggested^{2,6} that these collagen genes evolved from an ancestral coding unit of 54 bp. Type IV collagen is a specific component of basement membranes and contains two genetically distinct polypeptides, the $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ chains⁷⁻⁹. It differs from the types I-III collagens in that it contains interruptions in the Gly-X-Y repeat sequence¹⁰⁻¹² and does not form ordered fibrillar structures^{7,13}. We have isolated complementary DNA and genomic clones for the mouse $\alpha 2(\text{IV})$ collagen chain and here characterize 64-, 123- and 182-bp exons in the Gly-X-Y coding domain of the gene. The data suggest that the $\alpha 2(\text{IV})$ collagen gene may have evolved differently from those encoding the fibrillar collagens.

From a mouse parietal endoderm (PE) cDNA library¹⁴, we isolated a cDNA clone which by messenger RNA hybrid selection and *in vitro* translation was shown to be specific for one of the type IV collagen chains. In Northern analysis (Fig. 1) this cDNA, PE69 (430 bp), hybridizes to a 6.4-kilobase (kb) RNA, distinct from the major 6.8-kb and minor 6.2-kb RNA species which hybridize with a mouse cDNA probe specific for $\alpha 1(\text{IV})$ collagen chain (refs 14, 15 and M.K. *et al.*, in preparation). The nucleotide sequence of PE69 cDNA was determined and an amino-acid sequence of 143 residues was derived from it which contained Gly-X-Y repeats with three interruptions (Fig. 2). This sequence overlapped and was nearly identical with the partial protein sequence available from mouse and human $\alpha 2(\text{IV})$ collagen chains¹⁶. From these results we conclude that the mouse PE69 cDNA, reported here, is specific for the $\alpha 2(\text{IV})$

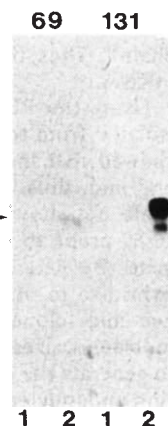
Received 4 March; accepted 16 July 1985.

1. Rudkin, G. T. & Stollar, B. D. *Nature* **265**, 472-473 (1977).
2. Langer-Safer, P. R., Levine, M. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4381-4385 (1982).
3. Bauman, J. G. J., Van der Ploeg, M. & Van Duijn, P. in *Investigative Microtechniques in Medicine and Biology* (eds Chayen, J. & Bitensky, L.) 41-87 (Dekker, New York, 1984).
4. Maniatis, T., Fritsch, E. F. & Maniatis, E. *Molecular Cloning* (Cold Spring Harbor, New York, 1982).
5. Baas, F. *et al. Hum. Genet.* **69**, 138-143 (1985).
6. Bauman, J. G. J., Wiegant, J., Borst, P. & Van Duijn, P. *Expl. Cell Res.* **138**, 485-490 (1980).
7. Renz, M. *EMBO J.* **2**, 817-822 (1983).
8. Shroyer, K. R. & Nakane, P. K. *J. Cell Biol.* **97**, 377a (1983).
9. Kriek, E., Miller, J. A., Juhl, U. & Miller, E. C. *Biochemistry* **6**, 177-182 (1967).
10. Landegent, J. E., Jansen in de Wal, N., Baan, R. A., Hoeijmakers, J. H. J. & Van der Ploeg, M. *Expl. Cell Res.* **153**, 61-72 (1984).
11. Van Ommen, G.-J. B. *et al. Nucleic Acids Res.* **11**, 2273-2285 (1983).
12. Baas, F., Bikker, H., Van Ommen, G.-J. B. & De Vijlder, J. J. M. *Hum. Genet.* **67**, 301-305 (1984).
13. Ploeg, J. S. in *Mononuclear Phagocytes in Immunity, Infection and Pathology* (ed. Van Furth, R.) 405-421 (Blackwell, Oxford, 1975).
14. Landegent, J. E., Jansen in de Wal, N., Ploeg, J. S. & Van der Ploeg, M. *J. Histochem. Cytochem.* (in the press).
15. Van der Ploeg, M. & Van Duijn, P. *Histochemistry* **62**, 227-232 (1979).
16. Dutrillaux, B. & Lejeune, J. C. *hebd. Séanc. Acad. Sci., Paris* **272**, 2638-2640 (1971).
17. Sumner, A. T., Evans, H. J. & Buckland, R. A. *Nature new Biol.* **232**, 31-32 (1971).
18. Gerhard, D. S., Kawasaki, E. S., Bancroft, F. C. & Szabo, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3755-3759 (1981).
19. Harper, M. E., Ullrich, A. & Saunders, G. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4458-4460 (1981).
20. Van Prooijen-Knegt, A. C. *et al. Expl. Cell Res.* **141**, 397-407 (1982).
21. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
22. Cox, K. H., DeLeon, D. V., Angerer, L. M. & Angerer, R. C. *Dev. Biol.* **101**, 485-502 (1984).
23. Heyting, C., Kroes, W. G. M., Kriek, E., Meyer, I. & Slater, R. *Acta histochem.* (in the press).
24. Morton, C. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2816-2820 (1984).

Fig. 1 Northern hybridization. Total cellular RNA (10 μg) of mouse embryo parietal endoderm cells (lane 2) and F9 embryonal carcinoma cells (lane 1) were run on an agarose-formaldehyde gel and transferred onto a nitrocellulose filter. One-half of the filter was hybridized with the PE69 cDNA probe and the other half with an $\alpha 1(\text{IV})$ collagen-chain-specific cDNA probe, PE131 (refs 14, 15). These probes hybridized to a 6.4-kb RNA (arrowhead) and a major 6.8-kb and minor 6.2-kb RNA, respectively, visible only in lane 2. As reported previously¹⁴, mouse parietal endoderm cells synthesize 100-fold more type IV collagen than do the F9 embryonal carcinoma cells. Size markers were mouse ribosomal RNAs²¹ visualized in the gel by ethidium bromide staining.

Methods. A cDNA expression library was constructed from total poly(A)⁺RNA of 13.5-day mouse embryo parietal endoderm (PE) cells by the double-linker method²². The double-stranded cDNA, with an *EcoRI* linker at the end corresponding to the 5' end of the mRNA and a *Sall* linker corresponding to the 3' end of the mRNA, was digested to completion with both *EcoRI* and *Sall* and size fractionated on a Sepharose 4B column. cDNAs ~500 bp and larger were then ligated to *EcoRI/Sall* double-cut vector pUC8 (ref. 23). Transformation of *Escherichia coli* strain DH-1 with recombinant plasmids was carried out as described previously²⁴. The bacteria were plated onto 82-mm nitrocellulose filters overlaid on ampicillin (50 $\mu\text{g ml}^{-1}$) plates to give 1,000-2,000 colonies per filter. For long-term storage at -70°C, colonies were replica plated onto nitrocellulose filters and replicas were grown on 5% glycerol plates²⁵. Duplicate filters of the PE cDNA expression library (~25,000 colonies) were screened with antibodies as described previously²² using a mixture of antisera against mouse laminin and ectactin²⁶ and mouse PYS-2 cells²⁷, a teratocarcinoma-derived PE cell line that produces large amounts of type IV collagen. After this 'shotgun' screening nine signal clones were selected for further analysis by Northern hybridization, mRNA hybrid selection and *in vitro* translation¹⁴. One clone (pPE69) was found to hybrid select mRNA for a polypeptide of 168,000 relative molecular mass, similar in size to one of the *in vitro* synthesized chains of mouse type IV collagen⁹. The mouse library was constructed in the hope that cDNA clones expressing sequences for basement membrane components or other abundant PE proteins could be identified directly by immunological screening. Indeed, in separate experiments, one cDNA clone expressing a fragment of mouse laminin B2 chain was isolated by this technique²⁸. However, isolation of the pPE69 clone from the PE cDNA expression library by antibody screening was only fortuitous. Sequence analysis of PE69 cDNA revealed, quite surprisingly, that the DNA strand corresponding to mRNA is in the wrong orientation for expression in the pUC8 vector (Fig. 2).

collagen chain. At present, the exact position of this sequence within the 1,740-amino-acid¹⁷ $\alpha 2(\text{IV})$ collagen chain is not known. However, the mouse protein sequence (Fig. 2) represents the N-terminus of pepsin fragment P2, of relative molecular mass 70,000, which is located near the N-terminal end of the



† Present address: European Molecular Biology Laboratory, 6900 Heidelberg, FRG.

Fig. 2 Nucleotide sequence and encoded amino-acid sequence of the mouse $\alpha 2(IV)$ collagen chain. First line: nucleotide sequence of the mouse PE69 cDNA. Second line: amino-acid sequence encoded by the cDNA. Third line: partial amino-acid sequence of the $\alpha 2$ -chain of type IV collagen from mouse tumour¹⁶. Fourth line: partial amino-acid sequence of the $\alpha 2$ -chain of type IV collagen from human placenta¹⁶. Symbols: —, amino-acid sequence which is the same as the cDNA-derived amino-acid sequence; x, unidentified amino-acid sequences marked by lines interrupt the Gly-X-Y repeat sequence. The numbering of nucleotides is for reference purposes only. On its 5' end, the PE69 cDNA sequence is flanked by a *SalI* octanucleotide linker sequence used for the cloning. The *EcoRI* recognition sequence on its 3' end, however, is present in the cDNA itself, as shown by sequence analysis of genomic DNA (Fig. 4). The PE69 cDNA-derived amino-acid sequence differs from the mouse protein sequence in five positions. Also, an additional Gly residue in the protein sequence¹⁶ is present between nucleotides 324 and 325. Because both strands of the cDNA were sequenced with compatible results and because we have carefully examined the sequencing gel data in the respective regions, we consider it unlikely that this variation is a result of an erroneous DNA sequence reading. In addition, the cDNA-derived sequence established that seven previously unidentified amino-acids were lysine. The lysyl residues in these positions are presumably modified post-translationally to glycosylated hydroxylysine and therefore not detected during amino-acid sequencing.

Methods. For sequencing according to the protocols of Maxam and Gilbert²⁹, pPE69 DNA was cut with *EcoRI* and the ends were labelled using T4 polynucleotide kinase and [γ -³²P]dATP. The end-labelled 430-bp cDNA was isolated by *SalI* digestion and polyacrylamide gel electrophoresis. To obtain the sequence of the complementary strand, pPE69 was first cut with *SalI*, end-labelled and then cut with *EcoRI*. For sequencing by the dideoxy method³⁰, using ³⁵S-dATP (ref. 31) and reverse transcriptase³² (Seikagaku America Co.), PE69 cDNA insert was released by *EcoRI*/*HindIII* digestion and cloned into similarly cut M13mp8 vector.

TCG ACC GAC CGG GGG CCC CCA GGA CCA CCT CCT CTC ATC CTG CCG GGG ATG	45
<i>SalI</i> linker Asp Arg Gly Pro Gly Pro Gly Pro Gly Pro Gly Pro Gly Pro Gly Met	
Pro - - - - - Val - - - - -	
AAG GAC ATC AAG GGA GAA AAG GGA GAT GAA GGA CCA ATG GGC CTG AAA GGG	96
Lys Asp Ile Lys Gly Glu Lys Gly Asp Glu Gly Pro Met Gly Leu Lys Gly	
- - - X - - - X - - - X - - - X - - -	
TAT CTG GGC TTA AAA GGC ATC CAA GGA ATC CCG GGA GTC CCC GGA GTG TCT	147
Tyr Leu Gly Leu Lys Gly Ile Gln Gly Met Pro Gly Val Pro Gly Val Ser	
- - - X - - - - - - - - - - - - - - - - -	
GGA TTC CCT GGG CTA CCT GGA AGG CCT GGC TTC ATC AAA GGA GTC AAG GGA	198
Gly Phe Pro Gly Leu Pro Gly Arg Pro Gly Phe Ile Lys Gly Val Lys Gly	
- - - - - X - - - - - X - - - - - X - - -	
GAC ATC GGA GTC CCT GGC ACA CCA GGC TTG CCG GGA TTC CCT GGC GTG TCT	249
Asp Ile Gly Val Pro Gly Thr Pro Gly Leu Pro Gly Phe Pro Gly Val Ser	
- -	
GGC CCT CCT GGA ATT ACC GGG TTT CCA GGA TTC ACA GGC AGC CCG GGC GAG	300
Gly Pro Pro Gly Thr Gly Thr Gly Phe Pro Gly Phe Thr Gly Ser Arg Gly Asp	
- - - - - Leu His - - - - - - - - - - - - - - -	
AAG GGT ACT CCA GGA GTA GCA GGA GTT TTT GGC GAG ACC GGC CCT ACT GGG	351
Lys Gly Thr Pro Gly Val Ala Gly Val Phe Gly Glu Thr Gly Pro Thr Gly	
X - - - - - Asp -Gly- Tyr - - - - - Ile - - - - - Ala - - -	
GAC TTT GGT GAC ATT GGG GAC ACT GTG GAC TTA CCA GGC GGC CCA GGC CTG	402
Asp Phe Gly Asp Ile Gly Asp Thr Val Asp Leu Pro Gly Gly Pro Gly Glu	
- -	
AAG GGG GAA CGG GGC ATC ACG GGA ATT C	
Lys Gly Glu Arg Gly Ile Thr Gly Ile	
X - - - - -	

chain¹⁶. Thus, the PE69 cDNA originated near the 5' end of the mRNA.

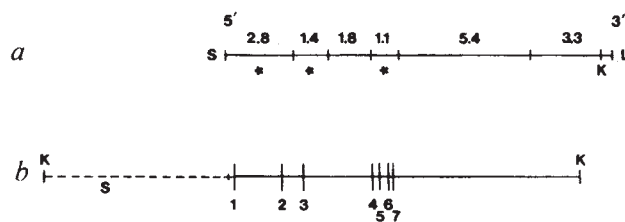
Using the PE69 probe, 1,100-bp cDNA clone (pE18) was isolated from the mouse PE cDNA library. Sequence analysis showed that this cDNA contains the PE69 sequence at its 3' end and, thus, that the PE18 cDNA extends 5' about 650 bp (data not shown). Mouse λ libraries were screened with the PE69 probe to isolate genomic clones for the $\alpha 2(IV)$ collagen gene. By heteroduplex analysis, PE18 cDNA was found to hybridize to at least seven small exons of 60–200 bp in one genomic clone, HA3. After examining 20 heteroduplex molecules, sizes of the exon and intron sequences were measured to generate the exon map shown in Fig. 3. In the *EcoRI* restriction endonuclease map of this genomic clone (Fig. 3), only the 1.1-kb fragment hybridized to the PE69 probe, whereas the PE18 probe, a 5' probe (see above), hybridized to the 1.1-, 1.4- and 2.8-kb fragments and established the 5' to 3' orientation of the genomic clone. As shown (Fig. 3), alignment of the two maps

suggests that four of the exons detected by heteroduplex analysis map close together within the 1.1-kb *EcoRI* fragment. When the nucleotide sequence of this fragment was determined, the sequences of three adjacent exons containing 64, 123 and 182 bp could be established (Fig. 4). Comparison with the cDNA sequence (Fig. 2) showed no nucleotide differences. These exon lengths agree well with those measured by electron microscopy and thus confirm the alignment of the restriction map and exon distribution map. The complete sequence of the fourth exon (at least 72 bp) remains to be determined because it extends beyond the 3' end of the 1.1-kb *EcoRI* fragment.

The 64-, 123- and 182-bp exons in the Gly-X-Y coding domain of the mouse $\alpha 2(IV)$ collagen gene, reported here, differ in several respects from those characterized so far in other types of collagen genes. First, in the fibrillar collagen genes (types I–III), most exons are 54 bp long and others are 45, 99, 108 and 162 bp long, that is, they all are multiples of 9 bp^{1–6}. This conservation of exon sizes has led to the suggestion^{2,6} that the

Fig. 3 Partial exon map. *a*, Mouse $\alpha 2(IV)$ collagen genomic clone HA3 was mapped with restriction endonucleases; the sites for *EcoRI* and *KpnI* (K) are shown. S and L refer to the short and long arms of the Charon 4A vector, respectively. Numbers denote the sizes in kilobases of the *EcoRI* fragments. The position of a 150-bp *EcoRI* fragment, present in this clone, remains to be determined. An 1,100-bp mouse $\alpha 2(IV)$ collagen cDNA probe, PE18, hybridized to the 2.8-, 1.4- and 1.1-kb fragments (asterisks). *b*, Heteroduplexes were formed between the PE18 cDNA and genomic clone HA3, analysed by electron microscopy, and the sizes of exons and introns measured. The results of these experiments are shown in the schematic drawing. PE18 cDNA hybridized to seven exons (vertical bars) of 60–200 bp. The exons are numbered 1–7 for reference purposes only. The horizontal line between the exons represents the approximate sizes of the introns. In addition, the following intriguing results were noted. Introns 2 and 3 (between exons 2 and 3, and 3 and 4, respectively) often hybridized to each other, indicating that they contain inverted repeat sequences. Also, in many heteroduplexes, the 3' genomic DNA hybridized to intron 2 and the 5' genomic DNA to intron 3.

Methods. A λ library of *HaeIII*/*AluI*-digested mouse liver DNA (P. Leder, personal communication) was screened³³ with the PE69 cDNA probe. One clone, HA3, containing ~16 kb mouse DNA was selected for further studies. With the PE69 probe, a 1,100-bp cDNA clone (pPE18) was isolated from the mouse PE cDNA library. Sequence analysis showed that the PE18 cDNA contained the PE69 sequence on its 3' end. Similar to the PE69 cDNA, PE18 cDNA also contained a *SalI* linker sequence on its 5' end and an *EcoRI* site on its 3' end. To determine the 5' and 3' orientation of the HA3 clone, *EcoRI* fragments were run on agarose gel and first hybridized³⁴ with the PE69 probe, then rehybridized with the PE18 probe. Heteroduplexes formed between the PE18 cDNA and the genomic clone HA3 were analysed by electron microscopy as described previously^{35,36}. For these experiments, to remove most of the vector sequences, HA3 DNA was cut with *KpnI*, which cleaves within the mouse DNA about 500 bp from its 3' end and removes the long arm of the Charon 4A vector. On its 5' end, 7.5 kb of short arm sequence remains after the *KpnI* digestion. The pPE18 cloned in pUC8 was cut with *DdeI* to produce a molecule in which the 1,100-bp cDNA remains flanked by vector sequences, 250 bp at its 5' end and 650 bp at its 3' end. To determine the 5' and 3' orientation of the HA3 clone by electron microscopy, heteroduplexes were formed using pPE18 DNA cut with *EcoRI*, which leaves the cDNA flanked by the 2.7-kb pUC8 vector sequence on its 5' end. In such heteroduplexes, a long DNA stretch was seen at the end with the large intron sequences.



Exon	Size(bp)	Sequence
4	123	ttccagcga AAC CGA GGA GAC CGG GGG CCC CCA GGA CCA CCT CCT CTC ATC CTG CGG GGG ATG Gly Lys Arg Gly Asp Arg Gly Pro Pro Gly Pro Pro Pro Leu Ile Leu Pro Gly Met AAG GAC ATC AAG GGA GAA AAG GGA GAT GAA GGA CCA ATG GGC CTG AAA GGG TAT CTG Lys Asp Ile Lys Gly Gly Lys Gly Asp Gly Pro Met Gly Leu Lys Gly Tyr Leu GGC TTA AAA Cgtagac Gly Leu Lys
5	182	ctccagcgc ATC CAA GGA ATG CCC GGA GTC CCC GGA GTG TCT GGA TTC CCG CTA CTT GGA Gly Ile Gln Gly Met Pro Gly Val Ser Gly Phe Pro Gly Leu Pro Gly Gly AGG CCT GGC TTC ATC AAA GGA GTC AAG GGA GAC ATC GGA CTC CTT GGC ACA GGC Arg Pro Gly Phe Ile Lys Gly Val Lys Gly Asp Ile Gly Val Pro Gly Thr Pro Gly TTG CCG GGA TCT CCT GGG GTG TCT GGC CCT CTT GGA ATT ACC GGG TTT CCA GGA TTC Leu Pro Gly Phe Pro Gly Val Ser Gly Pro Pro Gly Ile Thr Gly Phe Pro Gly Phe ACA GGC AGC CCGtagac Thr Gly Ser Arg
6	64	tcttagggc GAG AAG GGT ACT CCA GGA GTA GGA GTT TTT GGC GAG ACC GGC CCT ACT GGG Gly Glu Lys Gly Thr Pro Gly Val Ala Gly Val Phe Gly Glu Thr Gly Pro Thr Gly GAC TTT Cgttagat Asp Phe
7	7	ttccagcgt GAC ATT GGC GAC ACT GTG GAC TTA CCA GGG GGC CCA GGC CTG AAG GGG GAA CGG Gly Asp Ile Gly Asp Thr Val Asp Leu Pro Gly Gly Pro Gly Leu Lys Gly Glu Arg GGC ATC ACG GGA ATT C Gly Ile Thr Gly Ile

Fig. 4 Mouse $\alpha 2(\text{IV})$ collagen gene contains 64-, 123- and 182-bp exons. The sequences of three exons were determined from the nucleotide sequence of the 1.1-kb *EcoRI* fragment of the genomic clone HA3 (Fig. 3). Exons are numbered as in Fig. 3. Because of the nucleotide redundancy, boundaries of the intron between exons 6 and 7 were chosen to be consistent with consensus splice sites³⁷ (underlined). The sequence of exon 7 (at least 72 bp) extends beyond the 3' end of the 1.1-kb *EcoRI* fragment.

Methods. The genomic clone HA3 was digested with *EcoRI* and the fragments were ligated to *EcoRI*-cut and phosphatase-treated pUC13 vector (Pharmacia PL Biochemicals). After transformation of DH-1 cells, clones containing the 1.1-kb *EcoRI* fragment were isolated using the PE69 cDNA probe. This fragment was partially sequenced by the Maxam-Gilbert technique²⁹. For dideoxy sequencing, this fragment was cloned in both orientations into *EcoRI*-cut and phosphatase-treated M13mp18 vector³⁸, and after *PstI* digestion the 300- and 800-bp fragments were cloned into *EcoRI/PstI*-cut M13mp8 vector.

fibrillar collagen genes have evolved by amplification of an ancestral 54-bp exon unit embedded in intron sequences. Exon sequences other than 54 bp presumably were derived from later recombination events between the 54-bp exons^{2,6}. Second, in the fibrillar collagen genes, exons coding for the Gly-X-Y repeat sequence always start with an intact glycine codon. In this respect exons in the $\alpha 2(\text{IV})$ collagen gene are also different. Three out of the four exons characterized here start with a 2/3 intact glycine codon (Fig. 4). Third, in the fibrillar collagen genes, the third base in Gly and Pro codons is almost always U or C (ref. 1). However, such an unusual third-base preference is not seen in the $\alpha 2(\text{IV})$ collagen gene. Thus, in the 46 Gly codons examined (Fig. 4), the third-base frequency for G, A, U and C was 0.22, 0.43, 0.04 and 0.31, respectively, and in the 24 Pro codons, 0.08, 0.37, 0.42 and 0.13, respectively. These results suggest that the $\alpha 2(\text{IV})$ collagen gene is structurally different and may have evolved independently from the fibrillar collagen genes.

Similar results to those described here have also been obtained by analysis of mouse $\alpha 1(\text{IV})$ collagen gene (Y. Yamada *et al.*, personal communication). All the exons characterized so far are distinct from the exon sizes found in fibrillar collagen genes. In addition, other novel exon sizes have recently been described in the chicken $\alpha 1(\text{IX})$ and $\alpha 2(\text{IX})$ collagen genes¹⁸. Type IX collagen is very unusual in that it contains three triple-helical domains separated by non-collagenous domains. It is of interest that 5 out of the 11 exons characterized start with a 2/3 intact Gly codon similar to the type IV collagen exons reported here. Based on the exon structure, fibrillar collagen genes were proposed to belong in one class and type IX collagen genes to another separate class¹⁸. It remains to be seen whether this class should also contain type IV collagen genes. Finally, the invertebrate collagen genes from *Drosophila melanogaster*¹⁹ and *Caenorhabditis elegans*²⁰ provide yet another type of exon organization. The two small (1 kb) collagen genes isolated from *C. elegans* contain only one or two introns and the analysis of

the *Drosophila* collagen gene also shows clearly that it lacks 54-bp coding units. Like the vertebrate type IV and IX collagen genes, these genes code for proteins containing several interruptions in the Gly-X-Y repeat sequence.

We thank Phil Leder for providing the mouse genomic libraries and Richard Rickles, Karl Tryggvason, Francesco Ramirez, Michael Condon, D. J. Prockop and Brigid Hogan for help and advice. M.K. was initially supported by an EMBO fellowship and this work was supported by the NIH grants GM-34090 (M.K.) and AM32380 (F.R.).

Received 10 April; accepted 15 July 1985.

1. Tate, V., Finer, M., Boedtker, H. & Doty, P. *Cold Spring Harb. Symp. quant. Biol.* **47**, 1039-1049 (1982).
2. Yamada, Y. *et al. Cell* **22**, 887-892 (1980).
3. Sandell, L. J., Yamada, Y., Dorman, A. & Upholt, W. B. *J. biol. Chem.* **258**, 11617-11621 (1983).
4. Dickson, L. A. *et al. J. biol. Chem.* **256**, 8407-8415 (1981).
5. Vogeli, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5334-5338 (1981).
6. Yamada, Y., Liao, G., Mudryj, M., Obici, S. & de Crombrughe, G. *Nature* **310**, 333-337 (1984).
7. Timpl, R. & Martin, G. R. in *Immunocytochemistry of the Extracellular Matrix* Vol. 2, 119-150 (CRC Press, Boca Raton, 1982).
8. Crouch, E., Sage, H. & Bornstein, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 745-749 (1980).
9. Kurkinen, M., Foster, L., Barlow, D. P. & Hogan, B. L. M. *J. biol. Chem.* **257**, 15151-15155 (1982).
10. Schupann, D. *et al. Biochem. J.* **220**, 227-233 (1984).
11. Babel, W. & Glanville, R. W. *Eur. J. Biochem.* **143**, 545-556 (1984).
12. Schupann, D., Glanville, R. W. & Timpl, R. *Eur. J. Biochem.* **123**, 505-512 (1982).
13. Kefalides, N. A., Alper, R. & Clark, C. C. *Int. Rev. Cytol.* **61**, 167-228 (1979).
14. Kurkinen, M. *et al. Nucleic Acids Res.* **11**, 6199-6209 (1983).
15. Solomon, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3330-3334 (1985).
16. Voss, T. thesis, Univ. Munich (1983).
17. Dean, D. C., Barr, J. F., Freytag, J. W. & Hudson, B. G. *J. biol. Chem.* **254**, 590-596 (1983).
18. Lozano, G., Ninomiya, Y., Thompson, H. & Olsen, B. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4050-4054 (1985).
19. Monson, J. M., Natzie, J., Friedman, J. & McCarthy, B. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1761-1765 (1982).
20. Kramer, J. M., Cox, G. N. & Hirsh, D. *Cell* **30**, 599-606 (1982).
21. Hassouna, N., Michot, B. & Bachelier, J.-P. *Nucleic Acids Res.* **12**, 3563-3583 (1984).
22. Helfman, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 31-35 (1983).
23. Viera, J. & Messing, J. *Gene* **19**, 259-268 (1982).
24. Hanahan, D. *J. molec. Biol.* **166**, 557-580 (1983).
25. Hanahan, D. & Meselson, M. *Gene* **10**, 63-67 (1980).
26. Cooper, A. R., Kurkinen, M., Taylor, A. & Hogan, B. L. M. *Eur. J. Biochem.* **119**, 189-197 (1981).
27. Wolfe, J., Mautner, V., Hogan, B. L. M. & Tilly, R. *Exp. Cell Res.* **118**, 63-71 (1979).
28. Barlow, D. P., Green, M. N., Kurkinen, M. & Hogan, B. L. M. *EMBO J.* **3**, 2355-2362 (1984).
29. Maxam, A. J. & Gilbert, W. *Methods. Enzym.* **65**, 499-560 (1980).
30. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
31. Biggin, M. D., Gibson, T. J. & Hong, G. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963-3965 (1983).
32. Karanthanis, S. *Focus* **4**, 6-7 (1982).
33. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
34. Purrello, M. & Balaz, I. *Analyt. Biochem.* **128**, 393-397 (1983).
35. Chow, L. T., Broker, T. R. & Lewis, J. B. *J. molec. Biol.* **134**, 264-303 (1979).
36. Lopata, M. A., Havercroft, J. C., Chow, L. T. & Cleveland, D. W. *Cell* **32**, 713-724 (1983).
37. Mount, S. M. *Nucleic Acids Res.* **10**, 459-472 (1982).
38. Norrander, J., Kempe, T. & Messing, J. *Gene* **26**, 101-106 (1983).

Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription *in vitro*

Katherine A. Jones & Robert Tjian

Department of Biochemistry, University of California at Berkeley, Berkeley, California 94720, USA

During a herpes simplex virus (HSV-1) lytic infection, three classes of viral genes are transcribed in a temporally regulated manner¹⁻⁵. The HSV-1 'immediate-early' (IE) promoter sequences contain multiple copies of a hexanucleotide sequence, GGGCGG, known as a GC box, and one or more copies of an 11-base pair (bp) conserved A+T-rich element, designated TAATGARAT. The TAATGARAT elements are thought to mediate the *trans*-activation of IE RNA synthesis by a virion-associated protein(s)⁶⁻¹⁴, and

the flanking G+C-rich sequences appear both to potentiate this induction and to direct IE promoter activity *in vivo*⁹⁻¹⁴. The similarity of the herpesvirus GC box repeats to those of the simian virus 40 (SV40) early promoter¹⁵⁻¹⁷ prompted the *in vitro* analysis of HSV IE transcription reported here. We show that the mammalian gene-specific transcription factor Sp1 (refs 18-20) binds to eight distinct regions of the HSV short terminal repeat and stimulates transcription 25-fold from the divergent IE-3 (ICP-4) and IE-4/5 promoters.

The long and short unique segments of the HSV-1 genome are flanked by paired terminal and inverted repeats that contain four of the five IE promoters (see map in Fig. 4). In the present study we have focused on an *in vitro* analysis of transcription directed by promoters from the short inverted and terminal repeats. The IE-4 and -5 RNAs initiate at equivalent positions within the short repeated segments, but splice into unique coding regions^{21,22} and thus represent distinct transcripts regulated by a duplicated promoter region (designated as the IE-4/5 promoter). The HSV short repeats also contain an origin of DNA replication (ORI; refs 23-25) and two copies of the IE-3 gene, which encodes the ICP-4 protein that has a key role in the induction of viral early and late transcription^{6,7,26-29}.

To facilitate the *in vitro* analysis of IE gene transcription, we used hybrid plasmid constructions in which the IE-4/5 or IE-3 promoter sequences were fused to the HSV thymidine kinase gene (*tk*) at the *tk* BglII (+50) site (Fig. 1). RNA synthesized *in vitro* from these hybrid constructs was assayed by primer extension with an end-labelled DNA primer complementary to *tk* RNA within the coding region. Optimal expression of the IE-3/*tk* fusion *in vitro* was found to require sequences from -41 to -110, upstream of the IE-3 TATA box at position -25 (Fig. 1a, lanes 1, 2). These *in vitro* results resemble those obtained on microinjection of IE-3/*tk* plasmids into *Xenopus* oocytes¹², where a promoter was mapped from position -50 to -108 relative to the RNA start site.

The IE-3 upstream region (-40 to -330) contains eight GC box repeats similar to the six tandem direct repeats that bind the Sp1 transcription factor and activate SV40 early RNA synthesis^{19,20}. In contrast to the SV40 promoter, however, the IE-3 repeats are spaced at greater intervals and are present in both orientations. To determine whether IE-3 promoter recognition is mediated by Sp1, the HeLa whole-cell transcription extract was fractionated¹⁸ and used to reconstitute IE-3 transcription *in vitro*. A combination of the partially purified RNA polymerase II and Sp2 general transcription fractions was insufficient to recognize the IE-3 promoter, whereas this combination is sufficient to direct transcription from other RNA polymerase II promoters¹⁸. Instead, IE transcription was found to be strongly dependent on the addition of the Sp1 fraction (Fig. 1a, lanes 6, 7). Sp1-dependent transcription was also observed with the -110 deletion mutant (Fig. 1a, lanes 4, 5), which contains only a single GC box. The transcriptional efficiency increases no more than threefold in the presence of additional sequences (compare lanes 5 and 7 of Fig. 1a), and this small effect was detected only in the Sp1-reconstituted reaction. A similar 3-5-fold enhancement of RNA synthesis by far-upstream sequences (-110 to -330) has also been observed in *Xenopus* microinjection experiments¹². The IE-4/5 promoter also contains multiple GC box repeats that are distinct from the IE-3 repeats in orientation and/or location relative to the RNA start site. Nevertheless, transcription from the IE-4/5 promoter was also found to be highly responsive to Sp1-containing fractions (Fig. 1b).

To determine whether this transcriptional activation is mediated by sequence-specific interactions of Sp1 with the promoter, as has been demonstrated previously for SV40 (ref. 19), the IE-3 promoter was tested for its ability to bind the Sp1 transcription factor in a DNase I footprint experiment³⁰. Isolated fragments, end-labelled on either DNA strand, were incubated with various amounts of partially purified Sp1, subjected to partial DNase I digestion and analysed as described in Fig. 2 legend. Five distinct regions (I-V) of nuclease protection were found within a 300-bp region upstream of the IE-3 RNA start site. Each binding site

covers 18-20 bp and is centred about a single GC box, independent of the GC box orientation. The first binding site (-71 to -90) lies within the IE-3 promoter sequences mapped *in vitro* (Fig. 1a) and *in vivo*⁸⁻¹³. The other sites are located in the far-upstream (-110/-330) region, which does not have a large effect on IE-3 transcription *in vitro* or in the frog oocyte¹², but

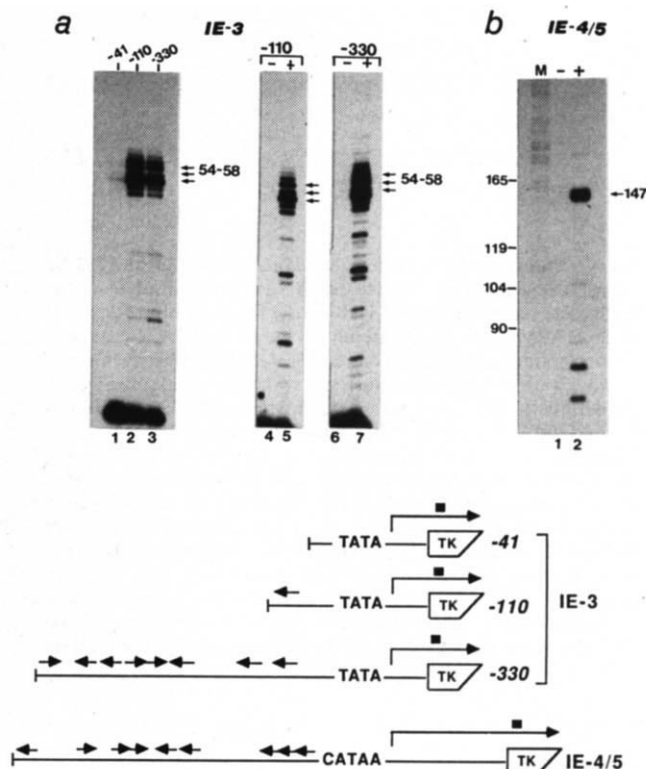


Fig. 1 Primer extension analysis of RNA synthesized *in vitro* from IE/*tk* fusion templates. The promoter regions of the hybrid IE/*tk* constructs are shown in the schematic drawing. The regulatory region of the IE-3 gene (-330 to +26) or the IE-4/5 gene (-420 to +116) was fused to the HSV thymidine kinase structural coding sequences at the BglII (+50) site as described elsewhere¹³. Note that *in vitro* transcription from these and similar hybrid constructs is regulated in a manner indistinguishable from that of unaltered IE genes^{13,14}. The 5' -110 IE-3 deletion mutant has also been described (ptkBX7; ref. 13). The -41 deletion mutant was constructed by preparative polyacrylamide gel isolation of a 220-bp fragment (from IE-3 -41 to *tk* +180) generated by *Nae*I digestion of the IE-3 -110 plasmid, and ligation into *Hinc*II-digested pUC9 vector³⁵. Short arrows indicate the relative location and orientation of the hexanucleotide repeats (GGGCGG); solid boxes show the hybridization position of the 24-nucleotide DNA primer on the IE/*tk* transcript. The primer extension protocol is described elsewhere³⁴. **a**, Lanes 1-3 show the extension products of RNA synthesized from 200 ng of circular plasmid DNA of the -41, -110 and -330 IE-3 templates respectively, each contained in a 50- μ l reaction mixture³⁴ containing 25 μ l (~50 μ g) of a 0.4 M KCl heparin-agarose step fraction (h.4) of the HeLa whole-cell extract. Subsequent chromatographic steps¹⁸ yielded the partially purified RNA polymerase II (CL. 225), and the Sp2 and Sp1 fractions used to reconstitute IE transcription, shown in the remaining panels. Comparison of lanes 4-7 reveals the transcriptional efficiency of the -110 and -330 IE-3 templates in the absence or presence of 8 μ l Sp1 (4 μ g; 12 μ l (36 μ g) RNA polymerase II, 4 μ l (2 μ g) Sp2 per 50- μ l reaction) (+ and - indicate the presence or absence of Sp1, respectively). Arrows indicate the fragment sizes expected for accurately initiated RNA transcripts. The transcription level in the Sp1-reconstituted reaction is 2-3 times higher than that observed in the h.4 fraction. Additional evidence that IE-3 RNA synthesis is Sp1-responsive derives from the observation that the promoter-specific factors required for IE-3 and SV40 early gene transcription in reconstituted reactions co-migrate on gel filtration columns used in either the conventional fractionation procedure or a different protocol (M. Briggs and R.T., unpublished) used to fractionate Sp1 derived from a nuclear extract (data not included). **b**, Transcription from a circular plasmid containing the IE-4/5 regulatory region in the absence (lane 1) or presence (lane 2) of 8 μ l (4 μ g) Sp1 (conditions as described for **a**); lane M, DNA size markers. All transcription reactions were incubated for 40 min at 30 °C, and RNA was prepared as described elsewhere³⁴. A small number of the IE-4/5 transcripts appear to initiate at reported upstream start sites³⁶.

enhances transformation 10–100-fold on transfection into mammalian cell lines^{13,31}.

Three regions of the IE-4/5 promoter are also recognized by Sp1 (Fig. 3). The far-upstream (III') site overlaps a single hexanucleotide whereas the larger regions (I' and II') overlap tandem GC box repeats. The IE-4/5 promoter has been mapped

in vivo to sequences within the I' site, and additional upstream sequences do not affect transcription in the absence of viral induction¹⁴. Interestingly, some of the IE promoter GC box elements are not recognized by Sp1 (open arrows in Figs 2, 3), which indicates that the GGGCGG sequence *per se* is not sufficient to specify a recognition site for Sp1. Comparison of the HSV immediate-early GC boxes with those of the SV40 early promoter suggests the following Sp1 recognition sequence: GGGCGG^{GC}_{AAT} (ref. 37).

Figure 4 shows a complete map of the Sp1 binding sites in the HSV short terminal repeat. Promoter activity *in vivo*^{12–14} and *in vitro* correlates with the Sp1 binding site proximal to each IE gene, and the distal Sp1 binding sites closely flank the

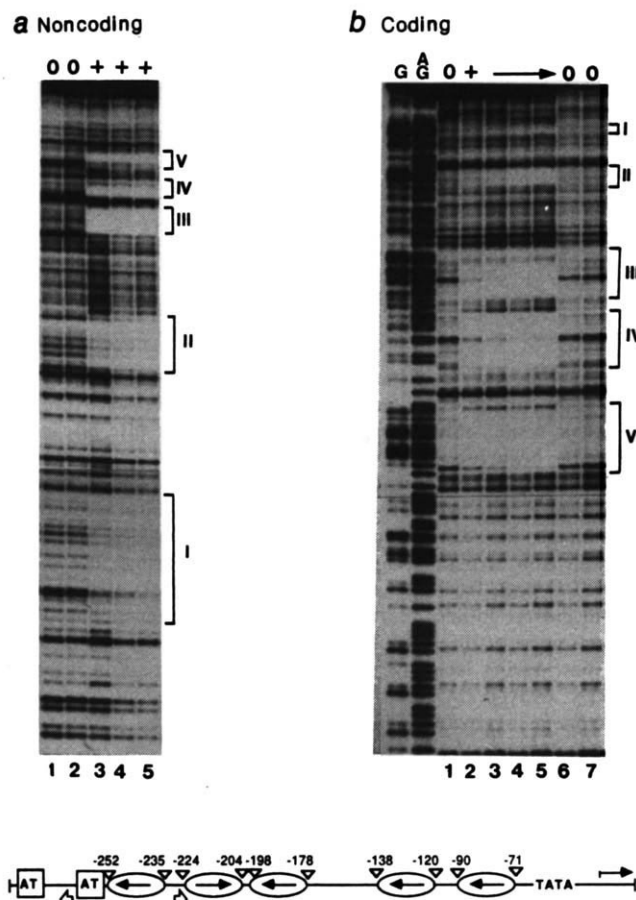


Fig. 2 DNase I footprint analysis of Sp1 binding sites in the HSV-1 IE-3 (ICP-4) promoter and upstream regulatory region. *a*, *b*, Autoradiographic exposures of 8% polyacrylamide sequencing gels used to analyse the partial DNase I digests of end-labelled IE-3 DNA fragments in the presence or absence of Sp1. *a*, Lanes 1, 2 show control DNase digestion patterns obtained in the absence of the Sp1 fraction. Lanes 3–5 show the DNase cleavage pattern in the presence of 10, 15 and 25 μ l respectively, of a 0.5 mg ml⁻¹ Sp1 fraction. Incubation and digestion conditions as in ref. 34. *b*, Lanes 1 and 6, 7 show the control reactions (no Sp1); lanes 2–5 show the DNase digestion patterns in the presence of 5, 10, 15 and 25 μ l Sp1 (0.5 mg ml⁻¹), respectively. The positions of these binding sites relative to the RNA start site are indicated on the map at the bottom of the figure. Solid arrows indicate the location and orientation of the GC box sequences central to each binding site; open arrows indicate two GC box hexanucleotides not recognized by Sp1 in this experiment; open boxes (labelled AT) show the positions of two TAATGARAT elements, one or both of which are required for viral activation of IE-3 transcription. Additional evidence that the binding sites represent interactions with Sp1 was provided by the observation that DNA fragments containing the SV40 21-bp repeats competed specifically for binding to the IE-3 sites (data not shown). The finding that a single Sp1 binding site covers only a 20-bp region is consistent with recent data which suggest that the upstream control region of the SV40 promoter contains multiple tandem Sp1 binding sites (D. Gidoni, J. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon and R.T., in preparation).

Methods. The plasmid used to generate the footprint probes was constructed as follows: the IE-330 recombinant plasmid (ptkBX8; ref. 13) was digested at the *Pvu*I site (–3, relative to the RNA start site), blunt-ended with *Bal*31 nuclease, digested with *Bam* endonuclease, and the 330-bp fragment was isolated by preparative gel electrophoresis and ligated into a *Bam*-*Hinc*II pUC9 vector³². DNA footprint probes were 5' end-labelled at the *Bam* site (*b*) or at the pUC9 vector *Hind*III site (labelling the opposite strand; *a*) using [γ -³²P]ATP and polynucleotide kinase, subsequently digested with *Hind*III or *Bam*, respectively, and the labelled fragments were isolated as described elsewhere³⁴.

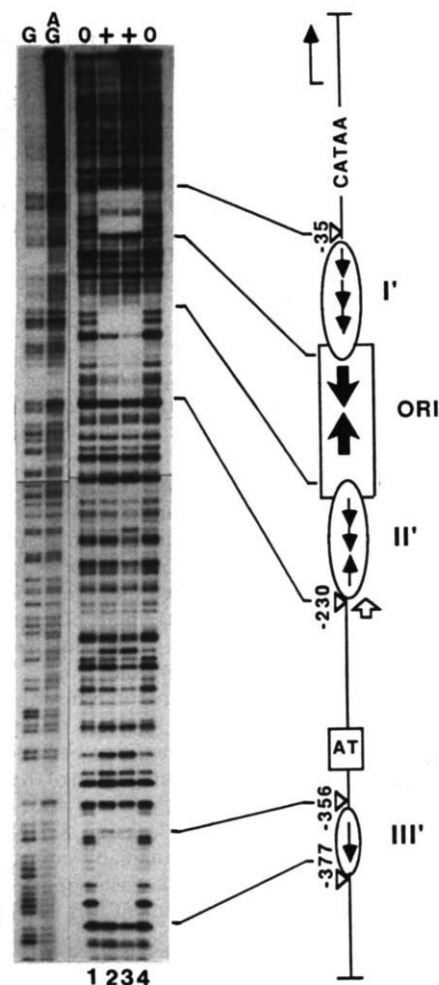


Fig. 3 DNase I footprint analysis of Sp1 binding sites in the HSV-1 IE-4/5 promoter and upstream regulatory region. The figure shows an autoradiographic exposure of a 5% polyacrylamide sequencing gel (acryl/bis = 29:1) used to analyse the partial DNase digests of end-labelled IE-4/5 DNA fragments. The IE-4/5 recombinant plasmid (ptkBX3; ref. 13) was 5' end-labelled at a *Bam*HI linker (–475 relative to the RNA start site), cut with *Sph*I and the 700-bp labelled fragment was isolated by gel electrophoresis. Lanes 1, 4, control DNase digestion pattern (in the absence of Sp1); lanes 2, 3, pattern obtained on incubation with 15 and 25 μ l, respectively, of a partially purified phosphocellulose Sp1 fraction¹⁸. G and A, G refer to guanine and purine sequence markers, respectively. The relative positions of the areas of DNA protected from digestion are indicated by ellipses on the map at the side. Small solid arrows indicate the position and orientation of GC box sequences present in each binding region; open arrow indicates a GC box element that is not recognized by Sp1. Also shown is the IE-4/5 TAATGARAT element (designated 'AT'), and the open box spanning sites I' and II' represents an origin (ORI) of DNA replication present in the short terminal repeat. Large arrows indicate a palindromic sequence within the ORI sequences.

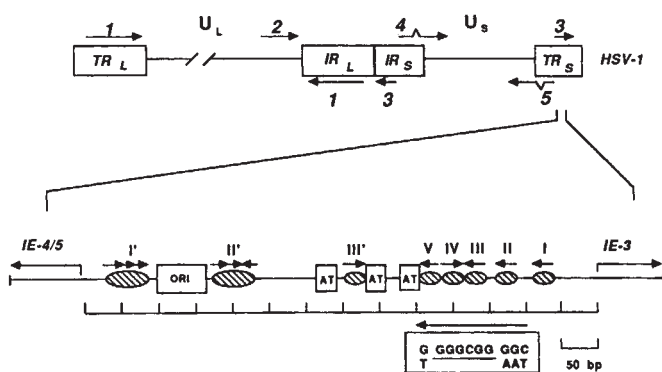


Fig. 4 Map of the Sp1 binding sites in the 700-bp region spanning the RNA initiation sites for the IE-3 (ICP-4) and IE-4/5 genes. Top, a schematic representation of the HSV-1 genome: TR and IR designate terminal and inverted repeats, respectively, and the L and S subscripts indicate long and short segments, respectively; U, unique sequences; the positions of the IE-1-5 transcripts are indicated by arrows. The portion of the small terminal repeat examined in the present study is expanded at the bottom. Sp1 binding sites are indicated by ellipses and the GC hexanucleotide repeats in each binding site are indicated by arrows. ORI, the palindromic sequence central to the origin of DNA replication; 'AT' boxes show the relative positions of the TAATGARAT elements implicated in the viral induction of IE transcription.

TAATGARAT elements ('AT' boxes) that have been implicated in *trans*-activation by the virion-associated protein. The sequences required for viral activation extend over 300 bp upstream of each gene, and appear to overlap only at the III' Sp1 site. Deletion of site III' sequences decreases the magnitude of the response of the IE-4/5 promoter to viral activation¹⁴, and suggests that Sp1 may potentiate this induction process *in vivo*. A complete definition of the role of Sp1 in IE transcription, however, clearly awaits a more detailed mutational analysis of these promoters. This extended arrangement of promoter and regulatory elements appears similar, in principle, to that observed for metallothionein and interferon genes^{32,33}, and may be a general feature of inducible eukaryotic transcription units. Finally, we note that the role of Sp1 in HSV transcription is apparently not limited to IE gene expression because Sp1 has also been implicated in transcriptional activation of the 'delayed-early' thymidine kinase gene³⁴.

We thank Jas Lang for pTKBX1, 3, 7 and 8 DNAs, Mike Briggs for nuclear extract Sp1 and our many colleagues who assisted with the manuscript. The work was supported by NCI grants to R.T., and K.A.J. was the recipient of a NIH postdoctoral fellowship.

Received 17 April; accepted 24 June 1985.

- Honess, W. R. & Roizman, B. J. *Virology* **14**, 8-19 (1974).
- Roizman, B., Kozak, M., Honess, R. M. & Hayward, G. *Cold Spring Harb. Symp. quant. Biol.* **39**, 687-701 (1975).
- Clements, J. B., Watson, R. J. & Wilkie, N. M. *Cell* **12**, 275-285 (1977).
- Jones, P. C. & Roizman, B. J. *Virology* **31**, 299-314 (1979).
- Anderson, K. P., Costa, R. H., Hallard, L. E. & Wagner, E. K. J. *Virology* **34**, 9-27 (1980).
- Preston, C. M. J. *Virology* **32**, 357-369 (1979).
- Dixon, R. A. F. & Schaffer, P. A. J. *Virology* **36**, 189-203 (1980).
- Post, L. E., Mackem, S. & Roizman, B. J. *Cell* **24**, 555-565 (1981).
- Mackem, S. & Roizman, B. J. *Virology* **43**, 1015-1023 (1982).
- Mackem, S. & Roizman, B. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4917-4921 (1982).
- Mackem, S. & Roizman, B. J. *Virology* **44**, 939-949 (1982).
- Cordingley, M. G., Campbell, M. E. M. & Preston, C. M. *Nucleic Acids Res.* **11**, 2347-2365 (1983).
- Lang, J. C., Spandidos, D. A. & Wilkie, N. M. *EMBO J.* **13**, 389-395 (1984).
- Preston, C. M., Cordingley, M. G. & Stow, N. D. J. *Virology* **50**, 708-716 (1984).
- Benoist, C. & Chambon, P. *Nature* **290**, 304-310 (1981).
- Myers, R. M., Rio, D. C., Robbins, A. K. & Tjian, R. *Cell* **25**, 373-384 (1981).
- Lebowitz, P. & Ghosh, P. K. J. *Virology* **41**, 449-461 (1982).
- Dynan, W. S. & Tjian, R. *Cell* **32**, 669-680 (1983).
- Dynan, W. S. & Tjian, R. *Cell* **35**, 79-87 (1983).
- Gidoni, D., Dynan, W. S. & Tjian, R. *Nature* **312**, 409-413 (1984).
- Murchie, M. J. & McGeoch, D. J. *J. gen. Virol.* **62**, 1-15 (1982).
- Watson, R. J., Sullivan, M. & Van de Woude, G. F. J. *Virology* **37**, 431-444 (1981).
- Stow, N. D. & McMonagle, E. C. *Virology* **130**, 427-438 (1983).
- Stow, N. D. *EMBO J.* **1**, 863-867 (1982).

- Mocarski, E. S. & Roizman, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5626-5630 (1982).
- Marsden, H. S., Crombie, I. K. & Subak-Sharpe, J. H. *J. gen. Virol.* **31**, 347-372 (1976).
- Watson, R. J. & Clements, J. B. *Virology* **91**, 364-379 (1978).
- Watson, R. J. & Clements, J. B. *Nature* **285**, 329-330 (1980).
- Everett, R. D. *EMBO J.* **3**, 3135-3141 (1984).
- Galas, D. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
- Preston, C. M. & Tannahill, D. *Virology* **137**, 439-444 (1984).
- Karin, M. *et al. Nature* **308**, 513-519 (1985).
- Goodbourn, S., Zinn, K. & Maniatis, T. *Cell* **41**, 509-520 (1985).
- Jones, K. A., Yamamoto, K. R. & Tjian, R. *Cell* (in the press).
- Messing, J. & Viera, J. *Gene* **19**, 269-276 (1982).
- Fontichiaro, K. L. E., Beck, T. W. & Milete, R. L. *J. Virol.* **53**, 235-242 (1985).
- Kadonaga, J. T., Jones, K. A. & Tjian, R. *Trends biochem. Sci.* (in the press).

Actomyosin structure in contracting muscle detected by rapid freezing

Shoichiro Tsukita & Masafumi Yano*

Department of Anatomy, Faculty of Medicine and * Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan

It is now widely accepted that the ATP-induced active sliding of adjacent thin and thick filaments mediated by myosin heads (cross-bridges) is responsible for muscle contraction. Despite intensive studies, the behaviour of the myosin heads during muscle contraction is still unclear. Recent progress in the rapid freezing electron microscope technique has greatly improved the temporal resolution of the images that can be obtained¹⁻³. Here, we report a new type of actomyosin structure captured by rapid freezing. We have analysed images from thin sections of freeze-substituted rabbit skeletal muscle rapidly frozen during isometric contraction. For comparison, we also studied relaxed and rigor muscles. Our results show that, during isometric contraction, most myosin heads are regularly arrayed along the helix of the actin filaments and that this actomyosin structure appears to be distinct from that observed in rigor muscle.

Glycerol-treated rabbit psoas muscle fibres were used. To measure the tension in the fibres up to the moment of freezing, the apparatus was modified (Fig. 1). The frozen muscles were freeze-substituted and embedded as described previously⁴. With this system, only the superficial layer (10-15 μ m deep) of the muscle was ultrastructurally well preserved. For detailed analysis at high magnification, we cut ultrathin sections <150 Å thick, which contain only a single layer of thin and thick filaments arranged in a hexagonal lattice⁵. As found previously by Huxley⁶, the longitudinal sections cut parallel to the crystallographic 1,1 planes of the lattice provided the most informative images of the crossbridges. The temporal resolution of these images is known to be at least 2 ms (ref. 1).

We first compared the morphology of thin filaments and crossbridges in the overlap region of relaxed and rigor (ATP-free) muscles (Fig. 2a, b). In relaxed muscles, the thin filaments appeared relatively smooth because they were decorated (labelled) by only a few crossbridges. N_c , which we define as the number of crossbridges per 0.5 μ m per thick filament in the 1,1 plane, was 8 ± 4 ($\pm$ s.d., $n = 20$) in relaxed muscles, although electron scattering caused by the embedding resin prevented our counting the exact number of crossbridges. In contrast, the thin filaments of rigor muscles were heavily decorated with many crossbridges ($N_c = 23 \pm 4$ s.d., $n = 20$), as a result becoming significantly thickened. These crossbridges showed the well-known 'arrowhead' configuration which is also characteristic of actin filaments decorated with an excess of isolated myosin heads (S1). In isometrically contracting muscles, we found that thin filaments are also extensively decorated with crossbridges ($N_c = 21 \pm 3$ s.d., $n = 20$) as abundantly as in rigor (Fig. 2c). The crossbridges in contracting muscles appeared to be angled slightly nearer to 90° than rigor crossbridges, although they still showed a preferred tilt pointing away from the Z band. At this stage, it is difficult to document further the difference in the tilting angle of crossbridges between rigor and contracting states, and careful

Fig. 1 Schematic diagram showing freezing of glycerol-treated muscle under conditions of isometric contraction. *a*, Freezing apparatus, modified from an Eiko apparatus (RF-10; Eiko Engineering Mito, Japan). *b*, Example of measurements of the isometric tension on an oscilloscope. *T*, tension; scale bar, 200 ms.

Methods. *a*, To measure isometric tension up to the moment of freezing, we designed a plunger (PL) to which a tension gauge (TG) was connected. We placed a specimen holder (SH), a tension gauge (AE801, Akers) and a small arm on the tip of the plunger. Glycerol was washed out for >4 h in rigor solution (4 mM MgCl_2 , 4 mM EGTA, 20 mM MOPS and 50 mM KCl, pH 7.0), then a small bundle of muscle fibres (M) was stretched over the specimen holder by fixing both ends to the tip of the tension gauge and the small arm using collodion. After the muscle on the tip of the plunger was relaxed in relaxation solution (4 mM ATP in rigor solution), it was transferred into contraction solution (4 mM CaCl_2 in relaxation solution). Tension development was recorded on a pen recorder. When the tension reached a maximum level, the plunger was inserted into the freezing apparatus and the isometric tension level (*T*) was continuously recorded by oscilloscope (*b*). The muscle was then rapidly frozen by slamming against a copper block (Cu) cooled to 4 K by liquid helium (He); the impact artefact and low temperature obscured recording of fibre tension (arrowhead). The copper block was moulded as shown in the diagram so that the tension gauge and the small arm did not collide directly with it. The frozen muscle was removed from the plunger together with the specimen holder and freeze-substituted as described in Fig. 2 legend. Because of the slow diffusion of the substrate into the muscle fibre, a concentration gradient of ATP may develop inside the fibre during contraction. To avoid this problem, we used an ATP-regeneration system, but found that the added proteins decreased the contrast of the freeze-substituted image. Therefore, because only the outer layer of the muscle ($\sim 10 \mu\text{m}$ deep) is analysed in the rapid freezing study, we have avoided using the ATP-regeneration system in this study, but used a sufficiently high concentration of ATP (4 mM).

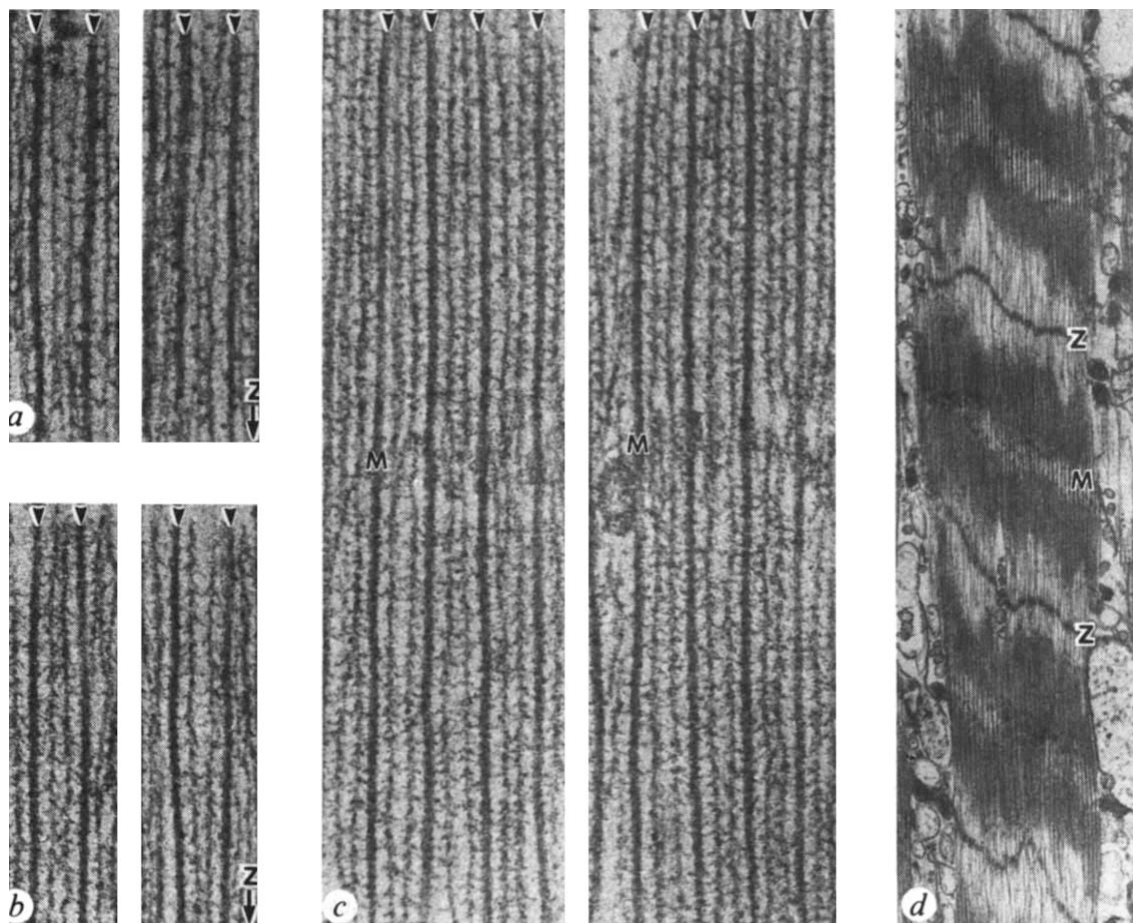
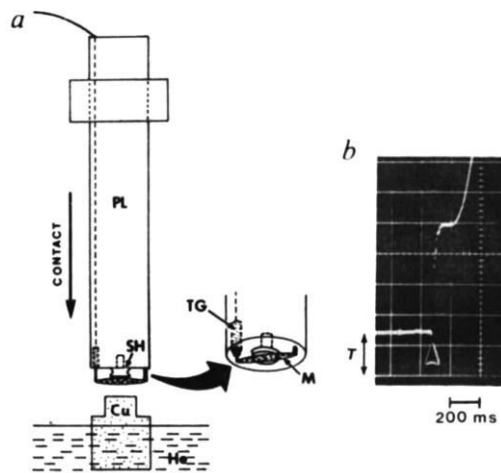


Fig. 2 Images of longitudinal thin sections of freeze-substituted muscles. *a*–*c*, Comparison of the images of the overlap region between relaxed (*a*), rigor (*b*) and contracting (*c*) states. In all experiments, images parallel to the 1,1 planes of the lattice, in which two thin filaments were observed between two thick filaments (arrowheads) were selected. These images confirmed that the thin sections used here contained only a single layer of thin and thick filaments. In the relaxed state (*a*), thin filaments contained less contrast, showing rather smooth contours with a few crossbridges. In sharp contrast, thin filaments in rigor muscle (*b*) appeared thick and were characterized by arrowheads pointing away from the Z bands (Z). The crossbridges had greatly increased in number. In isometrically contracting muscle (*c*), thin filaments also appeared thicker than those in relaxed muscle. Note the large number of crossbridges between the thin and thick filaments. $\times 97,750$. *d*, Low-power electron micrograph of freeze-substituted muscle rapidly frozen in conditions of isometric contraction. Thickness of section, 200–300 Å. The most striking feature of the contracting muscle is that the Z bands (Z) and the M bands (M) were seen as zig-zag lines. There were no detectable ice crystals. $\times 15,300$.

Methods. The rapidly frozen muscles were freeze-substituted for 2 days with absolute acetone containing 4% OsO_4 at -80°C . They were warmed by being transferred to a freezer held at -20°C where they were kept for 1 h and then stored at 4°C for 2 h. After being washed in absolute acetone, the samples were stained *en bloc* for 2 h at room temperature in absolute ethanol containing 5% uranyl acetate. They were then washed in absolute ethanol, then in propylene oxide and embedded in Epon 812. Ultrathin sections $<150 \text{ Å}$ thick were cut tangential to the metal contact surface and stained with 4% uranyl acetate and 0.5% lead citrate for 3 min each.

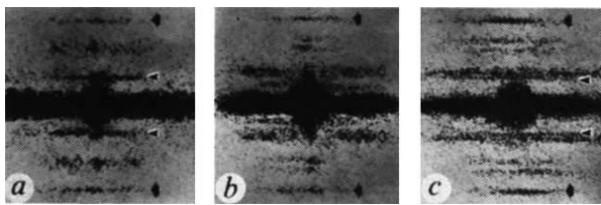


Fig. 3 Optical diffraction patterns of the whole A band from longitudinal thin-section images of freeze-substituted muscles. *a*, Relaxed state; *b*, rigor state; *c*, isometric contraction. The muscle showed characteristic diffraction patterns in each set of conditions. Black and white arrows represent the reflections with spacings of 143 Å (myosin subunit repeat) and 370 Å (crossover repeat of actin filament helix), respectively. The reflections with a spacing of 429 Å ($143 \text{ Å} \times 3$) are indicated by arrowheads. In isometric contraction, all sarcomeres examined ($n = 70$) showed the 370-Å layer line, whereas only 13% of the sarcomeres produced the 429-Å layer line. In addition to these major reflections, reflections with spacings of ~185 and 230 Å were clearly identified both in rigor and in isometric contraction, whereas in relaxation the so-called 'forbidden' reflection with a spacing of 215 Å was observed. The optical diffraction patterns in the rigor and relaxed states were mostly consistent with X-ray patterns of frog intact muscle, but our patterns were different in some respects. For example, in rigor state, some sarcomeres (6 of 71 sarcomeres; =9%) provided a layer line at 700–800 Å which was not reported in the X-ray pattern, and, in the relaxed state, the 429-Å layer line had a rather small radial extent compared with that seen in the X-ray diffraction. We suppose these differences may be caused by a lattice-sampling effect enhanced by the staining properties of the thin and thick filaments, but further studies are required to settle this question.

Methods. Using longitudinal thin sections ~500–700 Å thick, optical diffraction patterns were taken with an optical diffractometer (Chuo Seiki) from a whole A band. A rectangular mask was overprinted on the electron micrographs, according to a 'built-in method' developed by Hosaka and Hosoi¹¹. In relaxed muscle, the spacings of the outermost meridional reflection were ~139 Å. Considering that in the X-ray diagrams of the relaxed muscle the spacing of the major reflection is 143 Å, the specimens may have shrunk by ~3% during freeze-substitution and embedding and the values of the three major axial repeats detected in the relaxed muscles should be read as 143, 215 and 429 Å. All other spacings were corrected using this 143-Å reflection as a calibration factor.

quantitative analysis using computer image processing is needed to clarify this difference. In this study, we focused our attention to the axial periodicities of the crossbridges using optical diffraction analysis.

Using thicker sections, we obtained optical diffraction patterns from whole A bands (Fig. 3). In relaxed muscle, three meridional reflections with axial spacings of 143, 215 ± 3 and 427 ± 8 Å ($\pm$ s.d., $n = 67$ sarcomeres) were predominant. Both the 143- and 427-Å reflections are thought to be caused by myosin filament organization, whereas the 215 Å reflection may correspond to the so-called 'forbidden' reflection seen in the X-ray diagrams of resting muscle⁷. In the rigor state, in addition to the 143-Å reflection that might arise from myosin filaments, a prominent layer line was observed with a spacing of 376 ± 12 Å ($\pm$ s.d., $n = 71$ sarcomeres). This spacing (370 Å) seems to correspond to the crossover repeat of the actin filament helix. However, this 370-Å layer line in rigor muscles seems unlikely to arise only from the helical arrangement of the actin filament subunits, because no reflection with this spacing was detected in the diffraction patterns from relaxed muscle. Therefore, this strong 370-Å layer line should be attributed to the myosin heads arrayed along the helix of the actin filament in rigor, corresponding to the 'arrowhead' configurations seen in images from thin sections. In muscles frozen during isometric contraction, both the 143-Å reflection and the 370-Å layer line (380 ± 9 Å, $n = 70$ sarcomeres) were clearly identified in all sarcomeres examined ($n = 70$; 5 different samples), and the relative intensity of the 143-Å reflection to the 370-Å layer line appeared to increase compared with rigor muscles. The occurrence of the

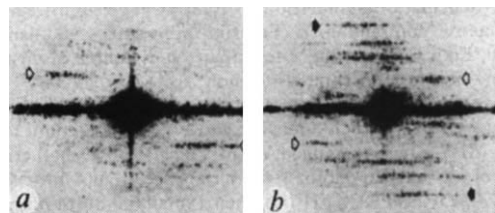


Fig. 4 Optical diffraction patterns of the single 'decorated' thin filament in the overlap region from rigor muscle (*a*) and isometrically contracting muscle (*b*). *a*, Rigor state. The layer line with the spacing of 370 Å (white arrow) was predominant. In this case, the meridional reflection with the spacings of 143 Å was hardly detected. *b*, Isometric contraction. Note that the 143 Å reflection (black arrow) has strong intensity. The 370 Å layer line (white arrow) was also detectable. Because the thickness of the sections used in this analysis was <150 Å, the optical diffraction patterns obtained from the single decorated thin filaments were occasionally one sided. These patterns were reproducibly obtained from thin filaments in different sarcomeres. A rectangular mask was overprinted on the electron micrographs to analyse single thin filaments ~0.6–0.8 µm long. The width of the mask was ~150 Å.

370-Å layer line clearly indicates that most of the myosin heads are arrayed along the actin filament during tension development. Time-resolved X-ray diagrams show that most myosin heads maintain the axial spacing of the myosin subunit repeat (143 Å) during isometric contraction^{7,8}. Taken together, these findings suggest that in contracting muscle most crossbridges retain the axial periodicity of their myosin filament origins (143 Å), are located along the actin filament helix and that the strong 143-Å reflection in the optical diffraction pattern of contracting muscles is attributed not only to myosin filament organization but also to the myosin heads labelling the actin filaments.

Optical diffraction patterns from single decorated thin filaments in contracting and rigor muscles show clearer evidence of the differences of their axial periodicities (Fig. 4). In contracting muscles a 143-Å reflection was reproducibly identifiable in addition to the 370-Å layer line, whereas in rigor muscle this 143-Å reflection was very weak or undetectable. Therefore, we conclude that, in the acto-S1 assembly of the contracting muscle, the myosin heads maintain the axial spacings of the myosin subunit repeat (143 Å), whereas in the acto-S1 assembly of rigor muscles most of the myosin heads are arranged according to the actin subunit repeat. These conclusions are mainly consistent with time-resolved X-ray diagrams obtained recently from isometrically contracting muscle^{7–10} and with our deep-etch replica images of contracting muscles (manuscript in preparation). Further analysis of the arrangement of myosin heads in the acto-S1 assembly in contracting muscles will lead to a better understanding of the molecular mechanism of muscle contraction.

We thank Drs H. Tanaka and Sa. Tsukita for collaboration in the rapid freezing experiments, Dr J. Usukura for technical advice, Dr J. Hosoi for collaboration in the optical diffraction experiments, Dr I. Matsubara for helpful discussions in the revision of this manuscript, and Drs H. Shimizu and H. Ishikawa for support. Rapid freezing with liquid helium was carried out in the Cryogenic Center, University of Tokyo. This study was supported in part by research grants from the Ministry of Education, Science and Culture, Japan.

Received 4 February; accepted 9 July 1985.

- Heuser, J. E. *et al.* *J. Cell Biol.* **81**, 275–300 (1979).
- Tsukita, S., Tsukita, S., Usukura, J. & Ishikawa, H. *J. Cell Biol.* **96**, 1480–1485 (1983).
- Heuser, J. E. *J. molec. Biol.* **169**, 123–154 (1983).
- Tsukita, S., Tsukita, S., Usukura, J. & Ishikawa, H. *Eur. J. Cell Biol.* **28**, 195–201 (1982).
- Reedy, M. K., Holmes, K. C. & Tregear, R. T. *Nature* **207**, 1276–1280 (1965).
- Huxley, H. E. *J. biophys. biochem. Cytol.* **3**, 631–673 (1957).
- Huxley, H. E., Faruqi, A. R. & Kress, M. *J. molec. Biol.* **158**, 637–684 (1982).
- Yagi, N., O'Brien, E. J. & Matsubara, I. *Biophys. J.* **33**, 121–138 (1981).
- Matsubara, I., Yagi, N., Miura, H., Ozeki, M. & Izumi, T. *Nature* **312**, 471–473 (1984).
- Huxley, H. E. *et al.* *J. molec. Biol.* **169**, 469–506 (1983).
- Hosaka, Y. & Hosoi, J. *J. ultrastruct. Res.* **84**, 140–150 (1983).

DNA makes protein makes money

from J. S. Emtage

The problems involved in producing unlimited quantities of hormones and similarly useful proteins by recombinant DNA techniques are fast being solved. With their solution comes the chance for those involved to reap almost equally unlimited financial rewards.

It is just twelve years since Stan Cohen, Herb Boyer and co-workers published the results of their work on transformation of *Escherichia coli* by R-factor (plasmid) DNA and on the construction of biologically functional bacterial plasmids *in vitro*. This work provided the cornerstone of recombinant DNA technology — the ability to cut and join DNA molecules very specifically and to introduce these molecules into *E. coli* in such a way that individual molecules, or clones, are selected. Of course, as with all major advances, other discoveries were soon made that expanded the power of the technology. In 1975 Ed Southern described a method for detecting specific sequences among DNA fragments separated by gel electrophoresis, the so-called Southern blot. We now have 'Northern' and 'Western' blot techniques for RNA and protein molecules respectively. Subsequently, in 1977, Maxam and Gilbert in Harvard and Sanger and colleagues in Cambridge independently developed methods for sequencing DNA. It is this combination of methods that has revolutionized molecular biology in the last decade. One practical result of this revolution is the renaissance of biotechnology with its potential applications in medicine, agriculture and industrial microbiology which a host of small companies hope to exploit.

It was the prospect of being able to produce unlimited quantities of proteins like insulin, growth hormones and the interferons that provided the driving force for the formation of these companies. At that time it was tacitly assumed that expression of heterologous genes in prokaryotic hosts like *E. coli* and the purification of the resulting protein would be simple. This has turned out not to be so and we now know that *E. coli* is not the perfect host for producing human or animal proteins. Many of them are found, after synthesis, in an insoluble, denatured form within the bacterium and tedious procedures involving chemicals like urea and guanidinium hydrochloride are required to render them soluble and active. *E. coli* has other problems too — frequently the initiation methionine is not removed from the protein and thus human growth hormone (hGH) for example is produced as MethGH. Moreover, *E. coli* does not add the carbohydrate side chains found on some eukaryotic proteins and, in situations where glycosylation is necessary for biological activity, this will clearly be a major

problem. Last but not least the complete removal of endotoxins from the product must be achieved during purification.

Because of these difficulties major efforts have been made to develop expression systems for yeast and mammalian cells. Yeast because of the existing knowledge of how to grow yeast on a large scale and because of its potential for secreting proteins and peptides into the medium and lack of toxins. Mammalian cells have been used because they will glycosylate proteins as well as secrete them. In both cases, secretion should result in a product with the correct N-terminal amino acid.

In the longer term the general opinion is that, for production, none of the above systems will predominate. Rather it will be 'horses for courses' and much will de-

pend on the final use of the protein, how much is required, the productivity of the different systems and the final cost. For some medical products where the volume required is low and a premium price can be charged, an animal cell system might be suitable. For other products like animal growth hormones where the predicted requirement is greater than 100 tonnes per year and the cost per dose must be low to make economic sense, then a route using a fermentable microorganism seems the only option at present.



As well as the progress discussed above on gene expression, all of which stems from an understanding at the molecular level of the interactions taking place during transcription and translation, recombinant DNA technology has had a fundamental effect on our ability to access

genes for proteins whose messenger RNAs are present in cells at low abundance. There are three methods that are commonly used at present. Genomic libraries made in phage vectors or complementary DNA libraries made in plasmids or phage vectors are screened using either a mixture of short oligonucleotides or a single long oligonucleotide as a probe. Obviously this approach is dependent on knowledge of some sequence of the protein of interest. In the second method cDNA libraries are made in an expression vector, such as the phage λ gt11, so that a fusion protein is produced consisting of the majority of β -galactosidase with the information encoded by the cDNA insert as a C-terminal fragment. Recombinant phage are then screened for the presence of antigenic fragments using an antibody to the protein.

The final method, unlike those above, does not require any detailed knowledge of the protein itself; all that is needed is a biological assay for the activity of the protein. In this system, libraries are made in *E. coli* in 'shuttle' vectors which are capable of high efficiency, transient gene expression in animal cells. Libraries are split into pools, DNA isolated from these pools and transfected into the animal cells. After incubation for 2–3 days medium from these cells is removed for bioassay. Once an active pool has been identified the specific clone in it is purified by sib selection. One or other of these procedures has been used in recent months in the cloning of the cDNA genes for protein such as tumour necrosis factor (TNF), colony stimulating factor (GM-CSF), human transforming factor- β as well as a number of receptor molecules.

Through genes and gene probes, recombinant DNA technology is also impacting on the diagnosis of (but not yet the treatment of) inherited diseases such as the haemoglobinopathies, haemophilia, phenylketonuria and Down's syndrome. The aim at present is to offer antenatal diagnosis and the possibility of a termination to pregnant women identified as at risk. This is done by isolating fetal cells by amniocentesis or from chorionic villi and culturing these to provide sufficient DNA for analysis. Gene probes are then used in Southern blots to show the presence or absence of the particular gene. This is convenient for gene deletions but what about point mutations? Until recently this has been a problem. However, work with

short oligonucleotide probes (19-mers) has shown that conditions can be defined where they will hybridize to normal but not to mutant alleles. Thus, provided the nucleotide sequence around the mutant site is known, then diagnosis using this approach should be possible.

Once a gene, or a gene family, has been cloned the present technology allows us to produce novel proteins either by 'shuffling' domains or by the more specific techniques of *in vitro* mutagenesis. Domain shuffling has been used mainly for the α -interferon family and really involves switching restriction fragments between two related genes. In this way a hybrid A-D α -interferon has been produced which shows different specificities from either parent interferon. *In vitro* mutagenesis however has far more potential and, if we can ever predict accurately what effect a particular amino acid change will have on the structure and function of a protein, the technique will have very wide application. It may, for example, be possible to increase the temperature stability of enzymes, to alter their pH profile or their substrate specificity. The most significant work done in this area is that of Fersht, Winter and colleagues on tyrosyl-tRNA synthetase. They have shown that single amino-acid substitutions can have dramatic effects on the catalytic parameters of the enzyme. This is certainly a good start along the road attempting to lay the ground rules for future protein engineering.

As well as being able to modify proteins by manipulating their genes it is also possible to modify pathways and so produce novel chemicals. An elegant example of this is the production of a hybrid antibiotic in *Streptomyces* by the transfer of biosynthetic genes between streptomycetes. Although the particular antibiotic produced in this experiment has no commercial value, just the demonstration of its production must be heartening to those interested in 'discovering' new antibiotics.

The most recent and interesting application of recombinant DNA technology is in the production of transgenic animals by injecting foreign DNA into the pronuclei or nuclei of eggs from superovulated animals. This produces offspring with the inserted DNA in terminally differentiated cell types. So far transgenic mice, rabbits, sheep and pigs have been produced, in most cases containing a fusion between the mouse metallothionein promoter/regulator region and either the rat or human growth hormone structural gene. Now that the technique has been established, however, it offers the possibility of investigating the mechanism of tissue-specific gene expression and maybe eventually of producing useful proteins in the milk of cows or in the eggs of chickens.

Big business for suppliers

The increasing number of laboratories engaged in recombinant DNA research is supporting a thriving industry.

- Instrument Research has introduced a high voltage pulse generator, the Model 958A, designed specifically for cell fusion systems. The instrument delivers a positive pulse adjustable from 0 to over 3.5kV, with pulse currents to 2A and peak powers over 7kW. This large voltage range permits adequate fusion gradients for large cells, and offers the further possibility of using series resistance in the cell circuit to approach a constant current characteristic. A digital pulse control module — 921A — permits direct setting of pulse width from 1 μ s to 99 μ s in 1 μ s steps, control of pulse period from 0.1 to 9.9s in 0.1s steps, and burst control of counted pulses from 1 through 99. An analog driver, Model 904A, offers continuous pulse control from less than 0.5 μ s up to seconds in 11 ranges, and also features triggered delay, burst, and single-shot modes. For maximum flexibility, the user can order both the analog and digital front ends, selecting either control via a rear panel connector. Applications include monoclonal antibody production (see M. Lo *et al. Nature* **310**, 792; 1984). *Reader Service No. 100.*

- Cambridge Biotechnology Laboratories, a division of Uniscience Limited, has added a further four restriction enzymes to the company's range. The new enzymes are *BanI*, *BclI*, *GspAII* and *ScaI*. *GspAII* is exclusive to CBL and recognizes the sequence TGCGCA. *Reader Service No. 101.*

- Transcriptaid RT from P & S Biochemicals of Liverpool is a new multi-component system for cDNA generation. By removing the need for the S_1 nuclease digestion step of ordinary procedures, Transcriptaid RT makes cDNA synthesis simpler and complete cDNA libraries can be constructed even from low abundance mRNA. In the new method, after the primary DNA strand has been synthesized with reverse transcriptase, the original mRNA strand is partially digested with RNase H. The remaining fragments then act as primers for synthesis of the secondary DNA strand using DNA polymerase 1. The reaction sequence is then completed by the removal of the non-transcribed 3' overhang with T4 DNA polymerase. A Transcriptaid RT kit contains five special reagents plus all the necessary buffers and nucleotides and control mRNA. The five reagents are used sequentially in a simple, step-by-step protocol. Each kit has enough material for

ten complete reactions starting with 5 μ g of mRNA.

Reader Service No. 102.

- A new range of *in vitro* RNA transcription products is now available from Pharmacia's Molecular Biology Division. MultiProbe pSPT 18 and pSPT 19 are two new SP6/T7 dual promoter plasmids containing the multiple cloning site from pUC 18 and 19 respectively, and together allow both strands of an inserted DNA to be transcribed. New FPLCpure T7 RNA Polymerase has exceptionally high functional purity, and has provided transcripts of up to 4.1 kilobases. Also now available are new RNaseStop= RNase/DNase-free bovine serum albumin and the ³mGpppN nucleotide caps necessary for translation of RNA transcripts. *Reader Service No. 103.*

- The lambda GT 11 sequencing primer is available as a kit from Clontech. A synthetic 15-mer adjacent to the cloning *EcoRI* site in lambda 11 is used for rapid DNA sequencing using the dideoxy method. This enables researchers to sequence clones identified by immunoscreening and other methods and eliminates the necessity to subclone in M13. The protocol is as simple as using M13 and yields sequence lengths comparable to M13. M13 primers are also available from Clontech. *Reader Service No. 104.*

- Bethesda Research Laboratories' biotinylated lambda DNA/*HindIII* fragments provide a convenient method for estimating the size of genomic DNA directly on nitrocellulose in Southern blot analysis. Following electrophoresis and transfer to a nitrocellulose filter, the fragments are visualized within 3h using a non-radioactive streptavidin/biotin detection method, such as BRL's DNA detection system. The size of the genomic DNA can be estimated by following three simple steps: (1) the biotinylated fragments and the biotin-labelled sample DNA are electrophoresed side-by-side on an agarose gel; (2) they are then transferred together onto a sheet of nitrocellulose by Southern blotting; and (3) the sample and the markers are visualized colorimetrically. *Reader Service No. 105.*

- Beta-lactamase screening filters are now available from Bio-Rad for the rapid detection of Amp^r recombinant clones. The filters provide a simple, rapid and reliable chromogenic assay for recombinant plasmids resulting from insertion of DNA into the ampicillin resistance gene of cloning vectors. After transfer to the sterile filters, recombinant (Amp^r) col-

onies appear blue, while non-recombinants (Amp^r) turn yellow within seconds. Colonies for further analysis may be picked directly from the filters, thus eliminating tedious replica-plating.

Reader Service No. 106.

• Boehringer Mannheim Biochemicals now offers two new restriction enzymes isolated from *Deinococcus radiophilus*, *DraII* and *DraIII*, and a protein-synthesis inhibitor, hygromycin B. *DraII* recognizes the sequence PuG↓GNCCPy and will produce fragments which have 5' cohesive termini. *DraIII* recognizes the sequence CACNNN↓GTG, generating fragments with 3' cohesive ends. Both enzymes are free of detectable nucleases and are ideal for use in all cloning and sequencing procedures.

Reader Service No. 107.

• For separation of biopolymers, Whatman now produces a comprehensive range of Advanced Ion Exchange Celluloses (AIEC) in microgranular (both pre-swollen and dry) and fibrous forms. Applications for these media include open column or batch techniques for the separation of proteins, nucleic acids, hormones, peptides, antibodies and other biological compounds.

Reader Service No. 108.

• A greater than 98% pure, high binding and precipitation capacity recombinant protein A is now available from Miles Scientific. Produced in *E. coli*, recombinant protein A is entirely free of staphylococcal toxins. The product's specific activity is equal to the most active protein A preparations extracted from *Staphylococcus aureus*. Over 95% of the recombinant protein A in all lots tested has been found to have active binding to an immobilized human IgG preparation. Recombinant protein A is supplied by Miles Scientific as a salt-free, lyophilized powder. The purity of every lot exceeds 98% as determined by SDS gel electrophoresis and HPLC analysis using a TSK-3000 column run in 0.2M NaPi at pH 6.8. Because of its high affinity for immunoglobulins of many species and its low level of nonspecific binding, recombinant protein A is useful in the study of cell surface antigens and receptors, and for detecting antibody-secreting cells.

Reader Service No. 109.

• Amersham Corporation is offering for the first time a slide chart which matches over 200 restriction enzyme recognition sequences with their restriction enzymes, isoschizomers, and enzymes which generate compatible ends. Developed by Amersham, the site selector measures only 4 inches × 9 inches folded, and concisely compiles information normally requiring cumbersome wall charts or multiple reference sources. The chart is available free of charge to workers in the field.

Reader Service No. 110.

• National Diagnostics has introduced a new Silver Stain Kit which is capable of detecting levels of polypeptide as low as 0.01 nanograms mm⁻² in electrophoretic

gels. It is suitable for both acrylamide and agarose gels and is extremely sensitive for the detection of single and double stranded DNA and RNA. National Diagnostics Silver Stain Kit is simple to use and is adequate for about 20 gels.

Reader Service No. 111.

• The new LKB Maxiphor gel casting and electrophoresis unit is designed for rapid preparative and analytical separations of DNA/RNA fragments with agarose submarine and bridge gels, including wedge-shaped field gradient gels. Even soft, low concentration agarose gels can be reliably cast and run with the new device. Maxiphor markedly simplifies the casting and handling of wedge gels, enhances resolution and improves linear frame reading. The unit accommodates horizontal agarose gels up to 10 mm in thickness, 15 cm in width and 25 cm in length. Up to 40 samples at a time can be separated with the double-comb technique. A complete agarose submarine system can be built up by adding the MacroDrive I power supply and MacroVue UV transilluminator to the basic Maxiphor unit.

Reader Service No. 112.

• Labquip's Caracol is a new DNA synthesizer for simultaneous multiple syntheses of oligomers. The Caracol principle is a vertical stack of PTFE disks containing ports which can be realigned by individual rotation. Reagents are delivered under a low pressure of inert gas to the upper 5 disks, which embody switching and mixing functions and pass reagents to the lower disks. Each of the disks has a reaction chamber containing a specially pre-treated paper, 6 mm in diameter, to which the first nucleotide is chemically coupled. Typically, 40 sequences (of oligonucleotide) are entered into the computer. All subsequent operations are controlled via an interface unit by a desk-top computer, which may if desired be several feet away. The package includes disk drive, program, computer, interface unit and cables.

Reader Service No. 113.

• The American Bionetics Inc. 2 × 2 Oligo Sizing Ladder is a set of electrophoretic standards for accurate length determination of oligonucleotides. Prepared from a mixture of purified synthetic DNA fragments ranging in size from 12 to 48 bases, these standards are useful for measuring oligonucleotides of up to 50 bases in length. The 2 × 2 Oligo Sizing Ladder may be used for length confirmation of synthetic DNA, as electrophoretic standards for DNA sequencing gels, and for electrophoretic determination of base composition of oligonucleotides. The ladder is provided lyophilized ready for reconstitution and end labelling with ³²P-ATP. Each tube contains sufficient material for 10 labelling reactions.

Reader Service No. 114.



ADVERTISEMENTS

PHORBOL
ESTERSPROTEIN KINASE C
ACTIVATOR —
PMA (TPA)

Cat. P1580. 5mg/ \$28.00



L.C. SERVICES CORPORATION

165 New Boston Street
Woburn, MA 01801
(617) 938-1700

Reader Service No.41

R-Phycoerythrin

HIGHLY PURIFIED
ECONOMICALLY PRICED

R-phycoerythrin is now available from

UPTON
BIOCHEMICALS

Please write for details to:—

Upton Biochemicals Ltd.,
20 Helena Road
Plaistow, London E13 0DU

Reader Service No.42

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID

Our new commercial LIPOSOMAT
guarantees rapid (5ml within 60 minutes)
preparation of uniformly sized

- UNILAMELLAR OR
- OLIGOLAMELLAR OR EVEN
- MULTILAMELLAR LIPOSOMES

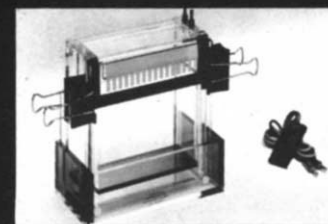
depending on the used experimental conditions. Liposome size is selectable between 25 and ca. 600 nm diameter up to 1000 nm for multilamellar liposomes. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. High encapsulation efficiency for hydrophilic and/or lipophilic agents. Lipid concentration can be up to 300 mg/ml. Lipid loss during preparation is extremely low: down to 0.1%.

BIANORM - Geräte

POB 126, D-8000 Munich 65, FRG

Reader Service No.39

THE LEADER HAS EMERGED
R. SHADEL, INC.
THE NEXT GENERATION IN ELECTROPHORESIS EQUIPMENT



YOUR ECONOMIC CHOICE FOR QUALITY

LET R. SHADEL, INC. SET THE STANDARD FOR
PROTEIN GEL ELECTROPHORESIS

- VERTICAL AND HORIZONTAL GEL UNITS
- ELECTROELUTION UNITS, STORAGE RACKS, RADIATION SHIELDS

THERE IS NO SUBSTITUTE FOR QUALITY
R. SHADEL, INC.

AVAILABLE EXCLUSIVELY FROM:

RAEL International

302 WEST 5th ST., SUITE 306
SAN PEDRO, CA 90731, U.S.A.

PHONE: (213) 547-5262
TELEX: 656 411 WU TELTEX LGB

Reader Service No.2

• Vector Laboratories is offering an applications note entitled 'Immunodetection of Antigens on Nitrocellulose Using the Vectastain ABC Kit'. The protocol provides clear stepwise instructions for the use of the Vectastain ABC immunoperoxidase system to detect antigens bound to nitrocellulose membranes. This protocol is applicable for electrophoretic or passive transfer systems including Western blots, dot-blots, or screening recombinant DNA libraries.

Reader Service No. 115.

• An Ultra Pure Acrylamide is now available in an electrophoresis grade from Schwarz/Mann Biotech. The acrylamide is manufactured under rigidly maintained standards to ensure a purity of greater than 99.9%. Specially controlled and tested, Ultra Pure Acrylamide can be successfully utilized in gel electrophoresis. Material is available now for shipment, packaged in 100 g, 250 g or 1 kg containers.

Reader Service No. 116.



Hoefer's rocking Deca-Probe.

• Especially suited to nuclear acid hybridizations and immunological screening, the new Hoefer PR 200 Deca-Probe is a 10-chamber micro-incubator manifold for the simultaneous probing of up to 10 membrane-bound samples previously fractionated on an acrylamide or agarose slab gel. Two solid acrylic plates fit together with an intact 15 × 15 cm transfer membrane between. Parallel troughs milled into one plate become the incubation chambers when the plates are joined. An O-ring gasket surrounding each chamber prevents leakage between them. A small port at one end of each 5 ml chamber allows introduction and removal of successive reagents. Rocking the unit ensures

adequate mixing with incubation volumes as small as 1 to 2 ml. Accessory 10-well combs align electrophoretic lanes to match the spacing of the Deca-Probe chambers.

Reader Service No. 117.

• Charles River Biotechnical Services, Inc. (CRBS) has developed a preliminary screening procedure for monoclonal antibody producing cell lines through Hybridoma Assessment and Evaluation (HAE) which includes the maintenance and expansion of hybridoma cells *in vitro*; storage of hybridoma cell lines in liquid nitrogen; MAP testing for murine viral contaminants; murine mycoplasma testing; pilot study ascites sample produced for approval; ascites yield determination; immunoglobulin yield determination; and production of one litre of raw ascites fluid. The tests, based on *in vivo* monoclonal production in mice, take approximately 12–15 weeks from the date CRBS receives the cells. Once the yield and viral status of a hybridoma are determined, further evaluation of that cell line is not required. Each aliquot of ascites fluid (after the initial litre) will be produced in approximately 6–8 weeks.

Reader Service No. 118.

• The 'shotgun' approach to the sequence analysis of large DNA molecules involves shearing into random fragments by sonication and then cloning these fragments into M13 bacteriophage vectors. A standard probe sonicator is not entirely suitable for this procedure due to the risk of DNA contamination. Life Science Laboratories of Luton now offer the Model W-375 sonicator with a special cup-horn probe to overcome this problem. The cup-horn is a flow-through device



The Life Sciences W-375 sonicator.

which uses water as both the transfer medium for sonication energy and sample cooling, allowing the DNA to be treated in sealed Sarstedt tubes in small volumes and without the risk of contamination. Distance from the vibrating tungsten probe, power control setting and time of treatment can all be adjusted to optimize results. The W-375 is a very powerful sonicator and breakage normally only requires 3 or 4 bursts of maximum energy of 30–40 seconds duration.

Reader Service No. 119.

• The Organics DNA Chemiprobe kit was a radioactive labelling system for DNA in which the DNA probe includes DNA chemically modified by a sulphone residue so that it becomes an immunogen.



Picogram quantities of DNA are detected by the Chemiprobe.

Monoclonal antibodies against sulphonic residues are conjugated to alkaline phosphatase. The presence of complementary DNA is detected by a colour reaction of a suitable enzyme substrate. The new probe detects picogram quantities of DNA. The system is applicable to nitrocellulose blots from plaques, colonies or gels.

Reader Service No. 120.

• The Calbiochem Hybridoma Screening Kit is a complete enzyme immunoassay system for the screening of mouse monoclonal antibody producing supernates. All necessary reagents are supplied to screen tissue culture supernates with minimum preparation. The assay system is based on the detection of antigen-specific antibodies by solid-phase absorption of the antigen followed by incubation of supernates and anti-mouse immunoglobulin peroxidase conjugate. Positive well colour development has been simplified using tableted *o*-phenylene diamine and a pre-formulated urea peroxidase substrate.

Reader Service No. 121.

• The Systec Microsyn-1440 interactive DNA synthesizer is a state-of-the-art synthesis system for the laboratory with a small to moderate oligonucleotide requirement. The Microsyn-1440 automates the repetitive synthesis protocol for each step, only requiring the operator to inject the base. The standard phosphite method produces coupling yields of 98% per step or better at a cycle time of less than 10 min per step. The instrument can synthesize using all current phosphate or phosphite methods, and at any scale from 0.5 to 10 μ M.

Reader Service No. 122.

• Photobiotin, a new reagent for the light-activated chemical coupling of biotin to DNA and RNA in the preparation of non-radioactive nucleic acid probes, is available from Biotechnology Research Enterprises S.A. Pty. Ltd. (BRESA). It reacts rapidly with both single- and double-stranded nucleic acids to give probes which are stable for many months. Such probes allow the sensitive colorimetric detection of target nucleic acids on either nitrocellulose or nylon membranes with avidin- or streptavidin-enzyme complexes. Potential applications include diagnosis of animal and plant diseases and *in situ* hybridization.

Reader Service No. 123.

ADVERTISEMENT

Automatic Tracking

Track that rat, insect, fish, limb, vehicle, crack, etc.

Tracking unit accepts input from video camera or recorder, and outputs the location of a high-contrast object in the TV image.

Can be interfaced to Apple II, BBC, etc.

HVS IMAGE ANALYSING LTD,
22 Cromwell Road, Kingston,
Surrey, England KT2 6RE
Tel: 01-546 9090



Reader Service
No.40